LEVEL II

# AIR FORCE INSTITUTE OF TECHNOLOGY

## AIR UNIVERSITY
## UNITED STATES AIR FORCE
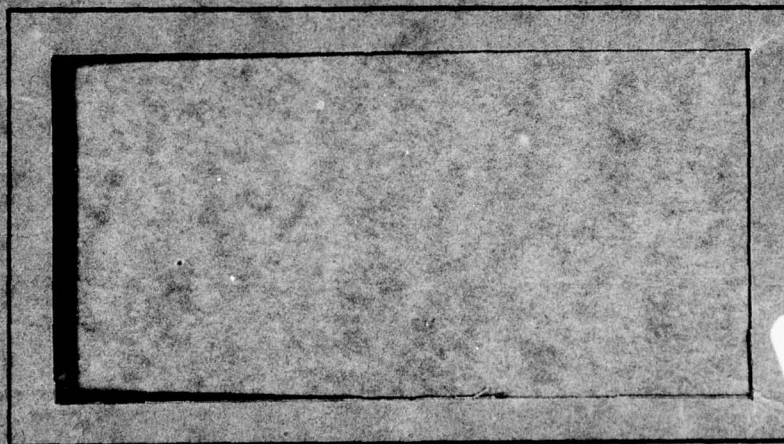
# SCHOOL OF ENGINEERING

## WRIGHT-PATTERSON AIR FORCE BASE, OHIO

78 07 07 010

LEVEL

AD A056516

D D C
JUL 20 1978
F

IDENTIFICATION OF PERIODICALLY

AMPLITUDE MODULATED TARGETS.

DISSERTATION

AFIT/DS/EE/78-2

CLAYTON V. STEWART
MAJOR            USAF

May 78

Doctoral Thesis,

137 p.

7622

01

78 07 07 010

012225

AFIT/DS/EE/78-2

IDENTIFICATION

OF PERIODICALLY AMPLITUDE MODULATED

TARGETS

DISSERTATION

Presented to the Faculty of the School of Engineering

of the Air Force Institute of Technology

Air University

in Partial Fulfillment of the

Requirements for the Degree of

Doctor of Philosophy

by

Clayton V. Stewart, B.S., M.S.

Major                               USAF

IDENTIFICATION

OF PERIODICALLY AMPLITUDE MODULATED

TARGETS

by

Clayton V. Stewart, B.S., M.S.

Major                          USAF

Approved:

_____                    23 May 78
Chairman

_____                    5/22/78

_____                    22 May 78

_____                    5/22/78

_____
George E. Orr                                              22 May 78

Accepted:

JSPrzemieniecki          23 May 1978

Dean, School of Engineering

## Acknowledgements

Many individuals have contributed greatly to this thesis. Thanks are due to Dr. Matthew Kabrisky, my principal advisor, for providing the inspiration for this work. I am also indebted to Maj Joe Carl and Capt George Orr for serving on my committee and for their critical reading of the dissertation. Special thanks must go to Dr. A. A. Ksienski, of Ohio State University, for his many timely suggestions and eminently practical advice. I am deeply grateful to Mehdi Shirazi and other personnel of the Radar Branch of the Air Force Avionics Laboratory for their outstanding cooperation in every way in the sponsorship of this project. Thanks are due to Mrs. Sandra Hastings for her excellent typing of the manuscript. Finally, I would like to express my most sincere gratitude to my dear family for the support they have provided during the years of my graduate education.

iii

# Contents

# List of Figures

# List of Tables

## Notation

| | | |
|---|---|---|
| a | – | Maximum radius; various other constants |
| a(t) | – | Amplitude modulating function (transmitted signal) |
| A | – | Transformation |
| $\alpha$ | – | Angular velocity |
| $\ast$ | – | Convolution operator (continuous or discrete) |
| b(t) | – | Amplitude modulating function (received signal) |
| B | – | Observation space |
| $\beta$ | – | Phase angle |
| c | – | Velocity of light |
| $c_{ij}$ | – | Cost of making decision i when j true |
| $\chi$ | – | Parameter space |
| $d(\cdot, \cdot)$ | – | Metric |
| $D_i(\cdot)$ | – | Linear discriminant function for class i |
| det | – | Determinant |
| $\delta(\cdot)$ | – | Delta function (Dirac or Kronecker) |
| $E_s(t)$ | – | Magnitude of scattered electric field |
| F | – | F-statistic |
| $F(\cdot)$ | – | Fourier transform (continuous or discrete) |
| $F_i$ | – | Magnitude of Fourier series component i |
| G | – | Pattern space |
| $\gamma$ | – | Linear rate of change of radian frequency |
| H | – | Feature space |
| $I_o(\cdot)$ | – | Modified zero order Bessel function |
| j | – | $+\sqrt{-1}$ |
| $\ell(\cdot)$ | – | Liklihood ratio |

| | | |
|---|---|---|
| $\lambda$ | - | Radar wavelength |
| $\Lambda(\cdot)$ | - | Wilks' lambda criterion |
| $M(t)$ | - | Complex modulating function |
| $\mu$ | - | Parameter vector |
| $n(t)$ | - | Noise |
| $N$ | - | A specific integer |
| $\omega$ | - | Radian frequency |
| $\omega_D$ | - | Doppler shift frequency |
| $\omega_i$ | - | Class i |
| $\omega_o$ | - | Radar carrier frequency |
| $\Omega$ | - | Finite set of class labels |
| $p(\cdot)$ | - | Probability density function |
| $P(\cdot)$ | - | Probability |
| $\phi$ | - | Aspect elevation angle |
| $r$ | - | Target range |
| $R$ | - | Bayes risk |
| $s(n)$ | - | Discrete time signal |
| $s(t)$ | - | Continuous time signal |
| $s_T(t)$ | - | Transmitted signal |
| $S(\omega)$ | - | Frequency-domain signal |
| $S_B$ | - | Between-class scatter matrix |
| $S_T$ | - | Total scatter matrix |
| $S_W$ | - | Within-class scatter matrix |
| $\sigma$ | - | Variance |
| $\Sigma$ | - | Covariance matrix |
| $t$ | - | Time |
| $T$ | - | A specific value of time |

## ABSTRACT

An algorithm that may be used for the classification of periodically amplitude modulated (PAM) targets is presented. The data base used to test the algorithm is derived from radar returns from vehicles moving at various velocities and aspect angles, but the techniques are applicable, as well, to other active wave devices such as sonar and laser. The received radar signal is considered to be a time series that is a function of target type, range, velocity, orientation and noise. Classification is implemented in the frequency domain; short-time spectra are computed using the Fast Fourier Transform (FFT). Features are extracted from the information bearing sidebands of the resulting spectra. The radar signatures are classified using both linear discriminant and nearest neighbor classifiers, and performance is presented for two, three, five, and six class cases using single and sequential looks. Probabilities of error of less than ten percent are achieved for five or fewer classes

# 1 Introduction

The purpose of this dissertation is to investigate techniques for the identification of radar targets that possess periodically amplitude modulated signatures. This introductory chapter includes a survey of automatic target identification with particular attention devoted to radar. A philosophical discussion of the general pattern recognition problem is presented followed by a procedure for the design of pattern recognition systems. The final section in the introduction is a summary of the ensuing chapters.

## Automatic Target Identification

Target identification is the act of assigning a label to the output of a specified sensor, or set of sensors, that has sampled an object of interest. Identification necessarily follows detection; the decision about the presence of a target must already have been made. In the following discussion and throughout this thesis, it is assumed that a target of interest is present and has been detected. Automatic identification, of course, implies that a machine, rather than a human observer, assigns the class label to the signal.

A hierarchy of levels of identification exist. If the universe of targets of interest is the set of all aircraft, the first division might be large aircraft vs. small aircraft. Among the class of large aircraft, there exist two natural classes: propeller-driven and jet. The set of large, jet aircraft may be further subgrouped into bomber and transport. The set of aircraft called B-52 belongs to the superset jet bombers. Various B-52 aircraft may be identified by model number, e.g. B-52B, B-52D, B-52H, etc. The most precise identification might, for example, assign the tail number to a particular B-52H observed. The designer of an automatic target identification system must consider the types of tar-

gets to be identified, the level to which they must be classified, and the types of sensors to be used.

Most target identification schemes described in the open literature are based on sensors that sample either acoustic or electromagnetic waves. The sensors may either be active or passive devices. An active device transmits a wave and receives the target's response to that wave. The target's electromagnetic or acoustic response is the input to the identification system. A passive device samples electromagnetic or acoustic emissions from the target and uses these emissions to classify the target.

Acoustic devices are used primarily underwater or on the surface of the earth. Underwater target identification by acoustic means generally implies sonar, either active or passive. Surface acoustic target identi-fication includes attempts to classify different types of vehicles by their acoustic emissions, e.g. Nichol (Ref 48) or Thomas (Ref 65). The identification of surface vehicles by analysis of their seismic signatures may be possible but is limited to a few hundred meters range (Ref 3:5) due to the rapid attenuation of such signals.

Electromagnetic target identification sensors may be classified by the portion of the spectrum in which they operate. In the higher fre-quencies one finds passive ultraviolet, optical, and infrared sensors. The outputs of such devices are typically images that may be analyzed by digital picture processing (Ref 59) or Fourier optics techniques (Ref 30). Pau (Ref 51) has described the use of a laser in a target recognition application. Radar operates in the microwave region of the electromagnetic spectrum. Radar target identification is the subject of this thesis.

Radar Target Identification. Since its development, radar has had

2

one glaring limitation: the inability to identify specific targets.

Radar operators and engineers have been striving to overcome this con-
straint for years (Ref 29). Skilled operators have achieved limited
success in target identification using conventional radars under certain
conditions. Azimuth resolution has been improved by using synthetic
aperture techniques where applicable. Range resolution has been in-
creased through pulse compression and wide bandwidth techniques. But
even if radar imagery can be made of optical quality, it would still
be desirable to perform target identification automatically in many
applications. In addition to robotic vision and perception, other
situations in which automatic identification would be desirable include
those in which the operator is relatively unskilled or must divide his
attention among many tasks.

Even when very high resolution systems are available, it may still
be more convenient to perform automatic identification in the signal
domain rather than the image domain. As will be demonstrated, under
certain conditions, features may be found in the signal domain that are
relatively invariant to target parameters such as viewing aspect. The
high resolution image of any interesting object, on the other hand, will
inevitably be aspect-dependent. This parametric dependence means that a
composite hypothesis testing procedure is required for identification,
rather than a simple hypothesis test that may be used for parametrically
independent signatures.

Numerous automatic radar target recognition techniques have been
proposed and tested with varying degrees of success. One of the most
fundamental approaches (Ref 41) attempts to characterize the target's
electromagnetic response by transmitting a set of harmonically related

frequencies in the Rayleigh region, where the dimensions of the scatterer are small compared with wavelength. Researchers at Ohio State University have had success identifying scale model aircraft and other objects using this approach. The obvious drawbacks to such a technique are the joint requirements of numerous, low frequency transmitters. Other researchers have selected features from the range trace of the target (Ref 58), necessitating a relatively high bandwidth. Numerous methods of radar target identification currently in use or under investigation are included in the survey by Nahin (Ref 47).

This thesis will present a method of classifying a specific set of radar targets, namely those that are periodically amplitude modulated. This class of targets includes many man-made, moving objects, since they are usually propelled by rotating structures that frequently have a radar cross section (RCS) that is a function of the rotation.

Radar Signals. An analysis of radar signals provides insight into what aspects of the signal may be useful for radar target identification. A monochromatic radar transmits a signal of the form

$$s_T(t) = a(t) \cos \omega_o t \qquad (1)$$

where $a(t)$ is a known amplitude modulating function and $\omega_o$ is the carrier frequency in radians per second. If the radar is of a continuous wave (CW) type, $a(t)$ is a constant. If the radar is a simple pulsed type, $a(t)$ is a positive constant for fixed intervals separated by longer fixed intervals where $a(t)$ is zero. For stationary radar, target, and clutter, the received signal is

$$s(t) = b(t) \cos(\omega_o t + \beta) + n(t) \qquad (2)$$

where $\beta$ is a random variable representing phase. Electrical noise is assumed to be an additive random process represented by $n(t)$. The

amplitude function b(t) is a random process that is a function of a(t), target range, clutter, and target RCS. The target RCS is, in turn, a function of target type, radar polarization, and viewing aspect.

If the radar remains stationary, but the target is moving at a fixed radial velocity, the form of the returned signal, neglecting clutter, is

$$s(t) = b(t) \cos [(\omega_o + \omega_D)t + \beta] + n(t) \tag{3}$$

The doppler shift is given by

$$\omega_D = \frac{4\pi v_R}{\lambda} \tag{4}$$

where $v_R$ is target radial velocity, and $\lambda$ is the radar wavelength. Neglecting noise, the positive frequency half of the amplitude spectrum of a particular realization of Eq (3) may be written as

$$|S(\omega)| = |B(\omega) * \delta(\omega - \omega_o - \omega_D)| \quad , \quad \omega \geq o$$

$$= |B(\omega - \omega_o - \omega_D)| \quad , \quad \omega \geq o \tag{5}$$

where $*$ represents convolution, $S(\omega)$ and $B(\omega)$ are the respective Fourier transforms of $s(t)$ and $b(t)$, and $\delta(\cdot)$ is a Dirac delta. If $b(t)$ is constant, the spectral representation will be discrete with a single spike at the frequency of the return. If, on the other hand, the RCS is fluctuating, $b(t)$ will be modulated resulting in a spreading of the spectrum. The amplitude spectra for the three types of targets discussed above are shown in Fig 1.

If prior knowledge exists about the form of $b(t)$, it may be used to aid in identifying the target being illuminated. If the doppler spread is sufficiently great, the nature of $B(\omega)$ may suggest the type of target. This is the essence of the radar target identification technique proposed in this thesis.

5
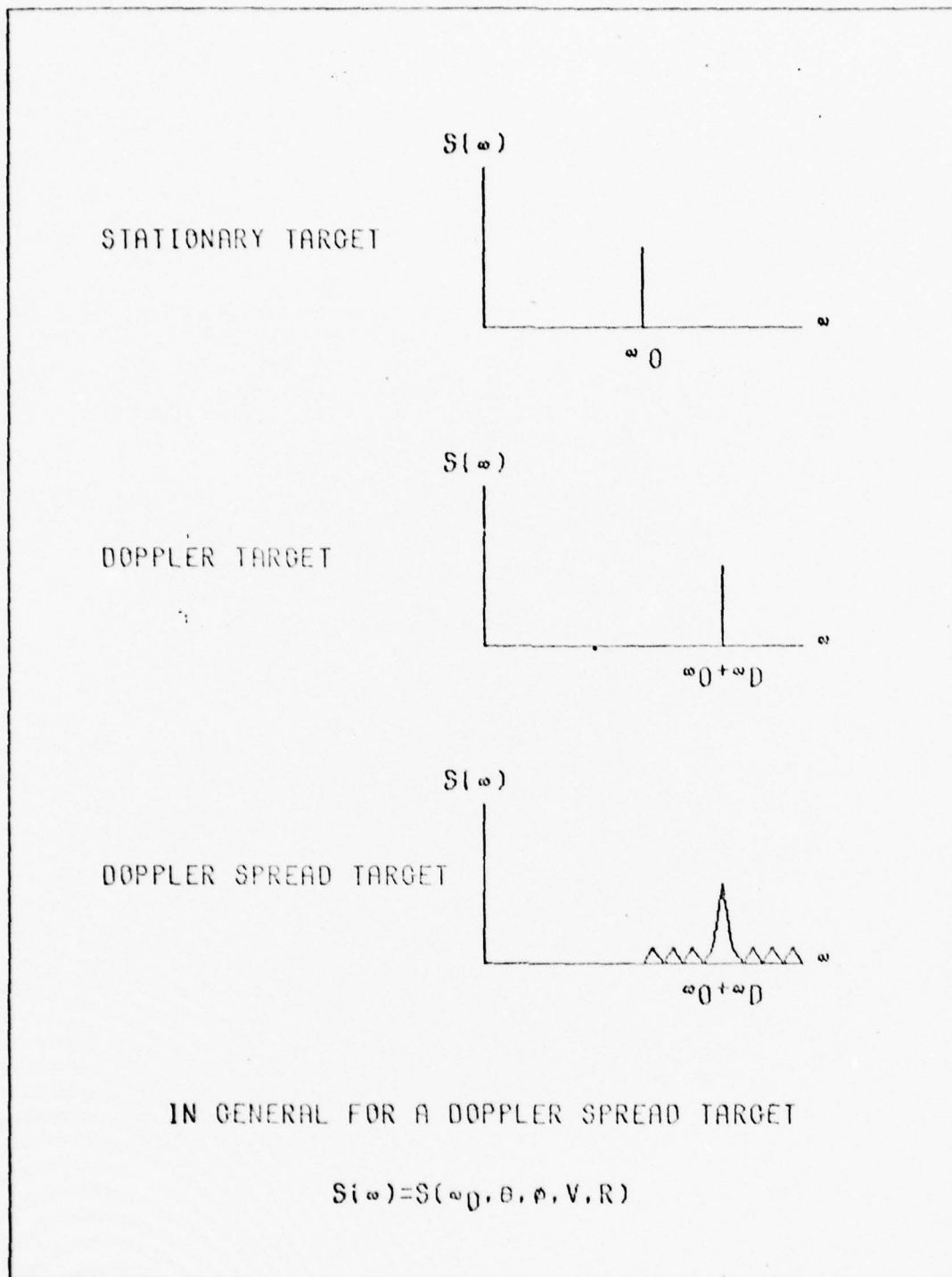
Fig 1. Typical Spectral Signatures of Radar Targets

## Pattern Recognition

For the past half of a century, numerous researchers and designers have been preoccupied with the idea of building a machine that can recognize patterns as humans do. Ullman (Ref 67) describes an optical character reader designed by Tauschek in 1929 that used a simple optical mask matching technique. Since that time, activity in the pattern recognition field has grown exponentially.

At first glance, there may appear to be little relationship between optical character recognition and radar target identification; however, both endeavors may be pursued within the mathematical framework of pattern recognition. Both the image of the alphabetical character and the returned radar signal may be regarded as mathematical functions, the former a function of two independent spatial variables and the latter a function of one independent temporal variable. The pattern recognition process consists of applying a series of transformations to the functions of interest. The determination of the best, or at least near best, transformations to apply is the art of pattern recognition.

Virtually any general discussion of pattern recognition contains an obligatory diagram such as Fig 2 (unless the author adopts a syntactic approach as espoused by Fu (Ref 24)). This figure, or some variant of it, has become the coat of arms of the pattern recognition researcher just as Shannon's (Ref 60) schematic diagram of a general communication system fills that role for the communications engineering community.

The real world, when considered as a vector space of potential measurements that could be made upon the universe, is of infinite dimensionality. People are continually sampling the world with sensors:
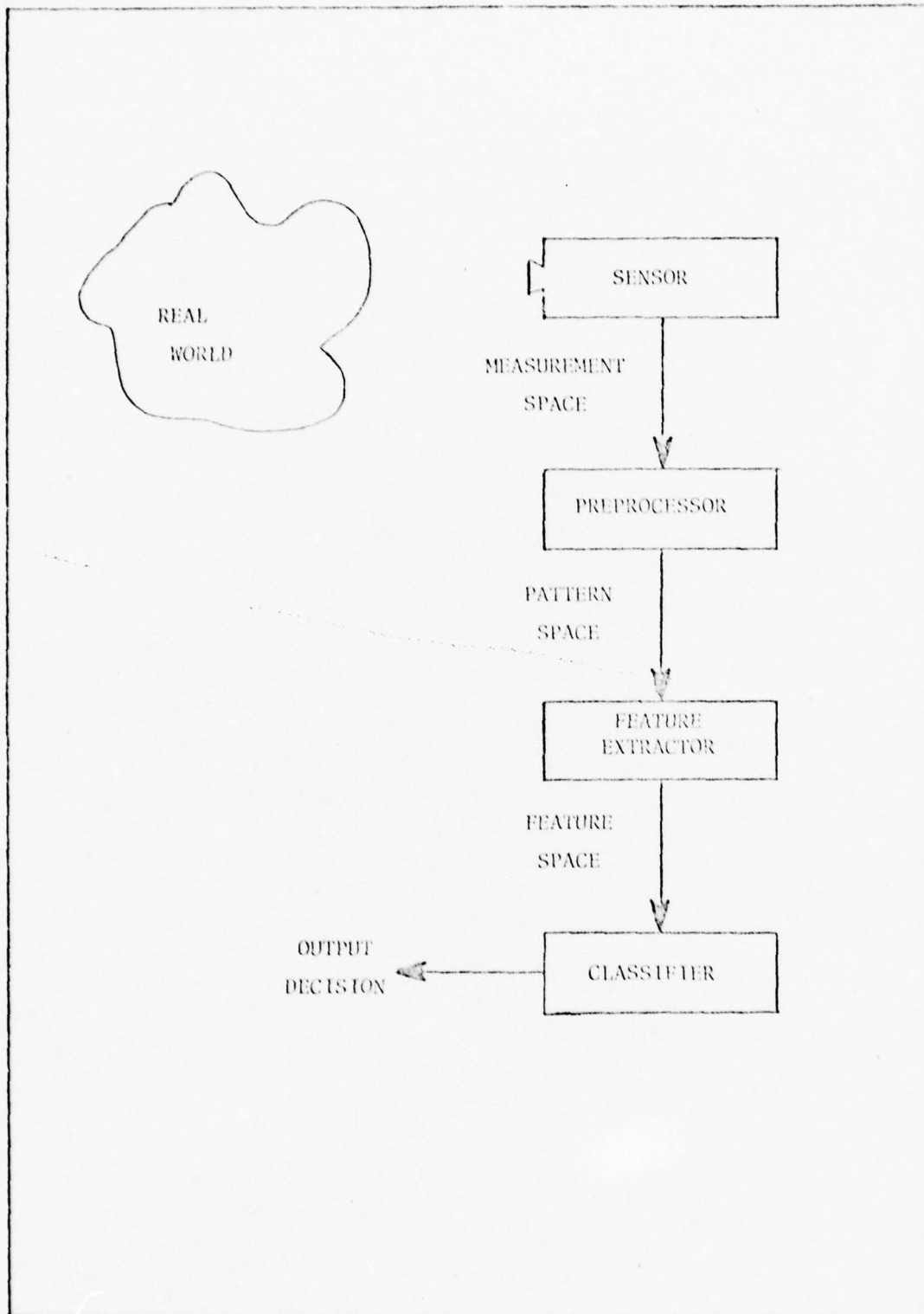
Fig 2. Canonical Pattern Recognition System

cameras, microphones, accelerometers, and radars to name a few. The output of the sensor may be discrete but it is more often analog, as is the case for the output of a radar receiver. In either case the sensor has reduced the dimensionality of the information at hand. The output of the sensor, considered as a function, resides in a vector space that is usually referred to as the measurement space. Physically, the sensor is a transducer that converts the desired physical quantities into a more convenient form, typically an electrical signal. Unfortunately, the sensor invariably injects into the desired signal some sort of noise. This noise may be due to any combination of measurement error, quantization error, thermal noise, external electromagnetic interference, leakage from the power supply or other source, intermodulation harmonics, or other signal distortion due to system nonlinearities.

The blocks following the sensor in Fig 2 are somewhat arbitrary, but they do seem to describe the three primary functions performed in a typical pattern recognition device. The distinctions between preprocessing, feature selection, and classification may not always be sharp. In fact, some authors do not honor preprocessing with a major block but relegate it to a subfunction under feature extraction.

As the block diagram implies, preprocessing is a transformation from the measurement space to the pattern space. Typically, the front end of the preprocessor consists of an analog-to-digital (A/D) converter, since the sensor is usually an intrinsically analog device and the remainder of the processing is done most conveniently in a digital system (considering contemporary technology). Another function of the preprocessor is the application of appropriate windows for data segmentation or presmoothing. Preprocessing may also include applying linear transformations to prewhiten or square up (i.e. normalize all measurements by their

9

respective standard deviations) the pattern space. Frequently, one of the most important functions of the preprocessor is to expand the data in a more convenient set of basis functions, e.g. via discrete Fourier transformation. Finally, the preprocessor may be used to filter out some of the noise and artifacts introduced by the sensor.

A brief digression in the form of a discussion of terminology appears necessary at this point. In the pattern recognition literature, "feature selection" and "feature extraction" are used synonomously to refer to four distinct transformations. Sometimes, the act of expanding the data in a new set of basis functions is referred to as feature selection; this transformation is included in preprocessing here. Next, in the design process a number of potential features are selected based on mathematical, physical, or statistical considerations; this process is designated design feature extraction by the present author. These features are evaluated and some useful subset is retained for classification. This process is termed feature selection here. Finally, in the on-line pattern recognition system, those features that have been selected in the design process must be extracted from the input data. This will be called feature extraction. It could certainly be argued that the terminology selected here is somewhat arbitrary, but at least it is fairly descriptive and does draw a distinction between the various feature selection/extraction types of processes.

In some applications the pattern space will be of adequately low dimensionality, and the class representations will be sufficiently separated that no further feature extraction is required. Such is not the case, however, for the problem at hand and probably seldom is for interesting pattern recognition problems. Thus the requirement for a feature extractor is indicated. Frequently, the pattern space will be

a Euclidean space of dimension between 100 and 1000. The feature extraction process is a dimensionality reducing transformation that should discard information that is common to the various classes and retain that information that bears the class discriminating capability. The feature space is typically of a dimensional order of ten.

Finally the feature vectors are input to the classifier which assigns class labels. Thus, the classifier, regardless of its implementation, parses the feature space and assigns labels to the various regions. Viewed abstractly, the pattern recognition system is a transformation from an infinite dimensional space, the space of all possible measurements on the universe, to a set, usually finite, of class labels.

In the following sections the various components of the pattern recognition system will be examined in greater detail, but the elaboration will be problem-specific. The reader who desires a more complete treatment of the general pattern recognition problem is referred to the texts by Andrews (Ref 1), Duda and Hart (Ref 18), Fukunaga (Ref 25), or Meisel (Ref 44). Each of the four have certain strengths and weaknesses. Andrews emphasizes feature extraction/selection and devotes considerable coverage to preprocessing transformations. Duda and Hart give good, even treatment to the subject including excellent bibliographies and historical presentations. Fukunaga is perhaps the most abstract and rigorous of the four, but he sometimes becomes so immersed in the formalisms that he loses sight of the goal of pattern recognition. Meisel is perhaps the most applications-oriented of the four. For a more complete critique of the texts, Cover's review (Ref 13) may be consulted. The reader who is more interested in a short survey of recent (1968-1974) work in pattern recognition will profit from Kanal's paper (Ref 39).

11

## Design of a Pattern Recognition System

The previous section discussed the implementation of a pattern recognition system, but the reader may well wonder how one arrives at this final design. The actual design process is frequently ignored in treatises on pattern recognition. The various transformations such as preprocessing, feature selection, and classification, are normally discussed in detail but little consideration is given to the actual integration of the various parts into a total system. Perhaps this omission is due to the fact that the design procedure is somewhat problem-specific. This section will attempt to generalize the pattern recognition design process as much as possible. The only constraints placed on the problem that is to be considered are that it is of a waveform recognition type where the signals of interest are generated by a physical process. Actually it would not be difficult to generalize to an imaging or higher dimensional type problem, since in general samples of those processes are simply functions of more independent variables. Also, systems other than physical, such as social, could be considered by simply generalizing the word "physics" to "system laws".

Two notable exceptions to the rule that little attention is paid to the design process are the book by Becker (Ref 2:71-85) and the thesis by Bouvier (Ref 6:85-88). Becker gives a "step by step description of the frequency of occurrence of binary words design procedure." It would seem particularly appropriate for Becker to do this, since his classification technique is radically different in its implementation from the more conventional schemes. Bouvier's design procedure, presented in his conclusions, is for seismic data, but its form is general enough for extension to other data bases. The greatest drawback to Bouvier's design procedure is that the first step is feature selection. He thus ignores

12

a great deal of the prefatory processing or implicitly incorporates it into feature selection.

The overall design process is depicted in schematic form in Fig 3. It may be noted that this figure bears some resemblance to the one that shows the canonical form of a pattern recognition system. Some of the blocks are rather similar, specifically those representing preprocessing, feature extraction, and classification. Because it is an adaptive, iterative procedure, the design process is frequently referred to as a learning or training process. This learning process requires a great deal of data processing, but once the final form of the system is determined, it usually will be implemented using a small device such as a microprocessor with its supporting components.

The first step in the design process, as Fig 3 indicates, is to investigate the underlying physics of the processes of interest. If the designer is not already well versed in this aspect of the problem, this investigation will necessitate a literature search and perhaps interviews with experts in the field. This step is one that many pattern recognition designers who are statistically or systems oriented appear loathe to pursue. A good understanding of the physics of the process will prevent problems from arising later in the design process such as extracting features that are based on artifacts in the data. Frequently, the underlying physics of the problem are not fully understood, and this lack of understanding is what may prompt a pattern recognition approach to the problem. In this case, the pattern recognition design procedure may well provide feedback in the form of greater insight into the underlying physics. In other cases the basic physical phenomenon is well understood, but there are so many sources of uncertainty that a pattern recognition

1. Investigate physics of processes of interest

2. Select an appropriate sensor

3. Consider the physics of the sensor

4. Generate a representative data base

5. Preprocess the data

6. Perform statistical analysis of the data

7. Simulation of the observed processes

8. Extract potentially useful features

9. Select "best" features

10. Select classifier

11. Evaluate performance

Performance satisfactory ?

No

Yes

Done

Fig 3. The Pattern Recognition Design Process

14

approach is specified.

After some understanding of the basic laws of the processes is gained, one must select an appropriate sensor, a device that will act as an interface between the environment and the pattern recognition device. Frequently, the type of sensor will be given; such is the case with the problem in this thesis. Numerous questions must be asked of any proposed sensor, such as does it have sufficient bandwidth and resolution? Must the sensor be a coherent device, i.e. must phase information be available? What sort of contamination of the received signal will the sensor introduce? These and other questions necessitate the feedback loop from block 3 back to block 2 in the diagram. Some of the questions about the sensor cannot be answered until the performance of the whole system has been evaluated. At that time, the designer may find that satisfactory performance simply can't be achieved with the specific sensor that has been chosen. On the other hand, the system design may indicate that some of the sensor complexity may be eliminated with a minimal sacrifice in overall performance. For example, coherence in a radar implies two video channels rather than one, which in turn means additional hardware and signal processing costs that may not be justifiable.

Following the selection of a candidate sensor, the designer must generate a representative data base. In practical terms, this means assembling representative samples of the objects to be identified, sense them under the various conditions under which they will most likely appear, and record their signatures. The collection of live data can be the most costly step in the design process; however, in some cases an excellent data base may already be available. The pattern recognition practicioner will seldom be an expert at taking electromagnetic, seismic,

or other physical measurements and must leave that work to those that are competent in the particular field of measurements. However, if the designer has a good understanding of the processes involved, he will be in a position to provide input as to the conditions under which the data must be taken. Also, if he has a good understanding of the measurement process, he will certainly have increased insight into the measured data. The designer's primary responsiblity in this area is to insure that the data collection is truly representative of the objects to be identified. For example, if three-dimensional objects that have aspect-dependent signatures are to be recognized, observations must be taken at various aspects to insure a representative data base.

Under some circumstances, the researcher may not have access to actual data, or it may not be feasible to gather it due to lack of the required resources. At times, such obstacles may be circumvented through physical or mathematical modeling. In his aircraft identification research, Ksienski (Ref 41) has used scale-model aircraft with conductive coatings and appropriately frequency-scaled radars. Lin and Richmond (Ref 50) have numerically computed the electromagnetic scattering signatures of aircraft based on wire grid models. The pattern recognition researcher must insure that such models accurately portray the processes of interest.

Preprocessing the data for the design procedure is similar to the corresponding step in the recognition system. Assuming that the classification will be done digitally, the data must be converted from an analog or instrumentation format to one that is computer compatible, and then filtered. Also, an appropriate pattern space must be chosen, one in which the physical peculiarities of the various classes are manifest.

16

After the data has been preprocessed, it should be subjected to a statistical analysis. Initially, the analysis will consist of the designer studying various plots of the data in order to correlate the representation of the data in the pattern space to the underlying physical processes. Then statistics of the process are computed, such as means and variances of potential features, as well as signal-to-noise ratios. Histograms, scatter plots, and cluster analysis routines may also be of value in this phase, and later when specific features are being chosen. One of the goals of the designer in this phase is to estimate the statistics of the data. It may be possible to apply probability density function estimation techniques such as Parzen windows (Ref 18:88-95).

Mathematical simulation of the measured processes forces the designer to quantify his observations. The modeling process may begin with block 1 and may be refined as the designer gains more insight. Simulation will also provide potential features to be used and may suggest the form of the classifier. For example, if all of the statistics of the data are known, a Bayes classifier is appropriate.

At this point, the designer will have in mind numerous features that bear class-discriminating information. General, as well as problem-specific, considerations in design feature extraction and feature selection will be discussed in greater detail in a subsequent chapter. The feedback loop from block 10 indicates that the designer must have in mind the type of classifier to be used when extracting and evaluating features.

After the features to be used are chosen and the form of the classifier has been specified, the overall performance of the pattern recognition system may be evaluated. The traditional figure of merit for a pattern recognition system is the mean probability of error. Toussaint (Ref 66)

17

has provided a definitive treatment of the estimation of the mean probability of error as well as an exhaustive bibliography of research on the subject. Shrihari (Ref 61) has defined classifier bias and has shown the importance of this measure in the multiobservation case. Relative economy of implementation in terms of quantity of hardware required and the time taken to make a decision are also relevant performance indicators. If a system is not feasible at present because of the quantity of data processing required, this may not be a serious long term limitation in view of the dynamics of large scale integrated circuit technology. It has been estimated that by 1985, it will be possible to build a hand-held calculator that possesses all of the computing power of the largest main frame computers of today (Ref 7). Performance evaluation will be considered in more detail in a subsequent chapter.

If the overall system performance is not satisfactory, the designer must return to some earlier phase and make appropriate revisions. The figure shows the feedback loop to the feature extraction stage. This is the most reasonable place to begin; however, the designer may ultimately be forced to return to any earlier phase including a re-examination of the basic physics of the processes of interest. In point of fact, although the feedback loops are not explicitly depicted in the diagram, the designer should feel free at any phase of the design process to return to some earlier stage to make refinements. There is, of course, no guarantee that this process will actually produce the desired classification performance in any arbitrary problem.

## Summary by Chapters

Throughout this discussion the underlying research question is how to distinguish between different radar targets that have periodically

amplitude modulated signatures.

In Chapter II the basic physical phenomena are investigated. The nature of the modulation is discussed, and the representation of modulated signatures in short-time Fourier space is developed. The manifestations of nonstationarities in the signatures are presented. The data bases, derived from returns from live radar targets, that are used to design and test a target classification algorithm are described.

In Chapter III, the identification problem is formulated in terms of optimal classification techniques. Discussions of Bayes classification, composite hypothesis testing, and sequential hypothesis testing are included. It is shown that, although the optimal techniques give insight into the nature of potential solutions to the problem, computational difficulties and lack of a complete physical understanding preclude their direct application.

Chapter IV presents a suboptimal frequency domain classifier. Preprocessing, feature extraction, and feature selection as performed on the design set data base are discussed. The signatures are identified using both linear discriminant functions and nearest neighbor classifiers.

Chapter V summarizes the conclusions to be drawn from this work. Areas of future research that pertain to the identification of radar targets that possess periodically amplitude modulated signatures are indicated.

## II  Periodically Amplitude Modulated Targets

This chapter discusses a class of radar targets that exhibit periodically amplitude modulated signatures. The physics underlying the modulation is presented, and its consequences in the frequency domain are examined. Finally, the data bases consisting of the returns from live radar targets, that are used in subsequent classification experiments, are described.

### The Physical Phenomenon

The high frequency radar echo from moving targets such as aircraft, ships, or vehicles is composed of the vector sum of a group of superimposed echo signals from the individual parts of the target. Nontranslational motion of the target and its associated components cause time varying fluctuations in the target RCS. Such fluctuations are manifest in the target signature in the form of a spectral spreading.

In classical radar detection, these fluctuations are considered to be target noise (Ref 19). A moving radar target that possesses an amplitude modulated signature is sometimes referred to as a doppler spread target, because of the resultant broadening of the frequency spectrum of the target. The optimum detector for a doppler spread target has previously been derived (Ref 70:357-375), but classification of different amplitude modulated targets is more difficult.

Dunn and Howard (Ref 19) have characterized these target fluctuations as amplitude noise and have segregated them into two types: low-frequency and high-frequency. The low-frequency fluctuations are due to random yaw, pitch, and roll motions of the target. This low-frequency variation, resulting in a broadening of the skin line of the target, was not used for target identification in this research. The skin line of a radar target is defined as that portion of the target's spectral signature that con-

tains the energy scattered by the skin of the target. Typically this energy is concentrated at a single frequency or over a very narrow range of frequencies as indicated in Fig 1 of the preceeding chapter. The high-frequency modulation consists of both random noise and periodic modulation. The random component results from skin vibration and random motion of target components. The periodic modulation is attributable to rapidly rotating parts of the target such as aircraft propellers, ship radar antennas, or vehicle running gear. The periodic RCS of a heavily-lugged, agricultural tractor tire, measured by Frost, is shown in Fig 4. The tire was mounted on a turntable, and its RCS was measured as a function of angle of rotation.

To understand the effect that a rotating structure has on the radar signature of a target, it is convenient to first consider a scatterer that has a simple geometry. Van Bladel (Ref 68) has presented a detailed exposition of electromagnetic fields in the presence of rotating scatterers. Chuang (Ref 11) has solved for the monostatic (i.e. source and observation points colocated) power spectrum of a rotating, polygonal cylinder in the high frequency region.

We may consider an infinitely long, conducting polygonal cylinder that is rotating at an angular velocity $\alpha$ about its longitudinal axis. The cylinder is in the far field of a radar of radian frequency $\omega_o$, with boresight normal to the cylinder's axis. Because of the rotation of the cylinder, the electromagnetic scattering is a periodic function of time with period T, where

$$T = \frac{2\pi}{N\alpha} \tag{6}$$

for an equilateral, N-sided polygonal cylinder. For $2\pi/T \ll \omega_o$, the magnitude of the backscattered field may be written as

$$E_s(t) = M(t)\exp(j\omega_o t) \tag{7}$$

21
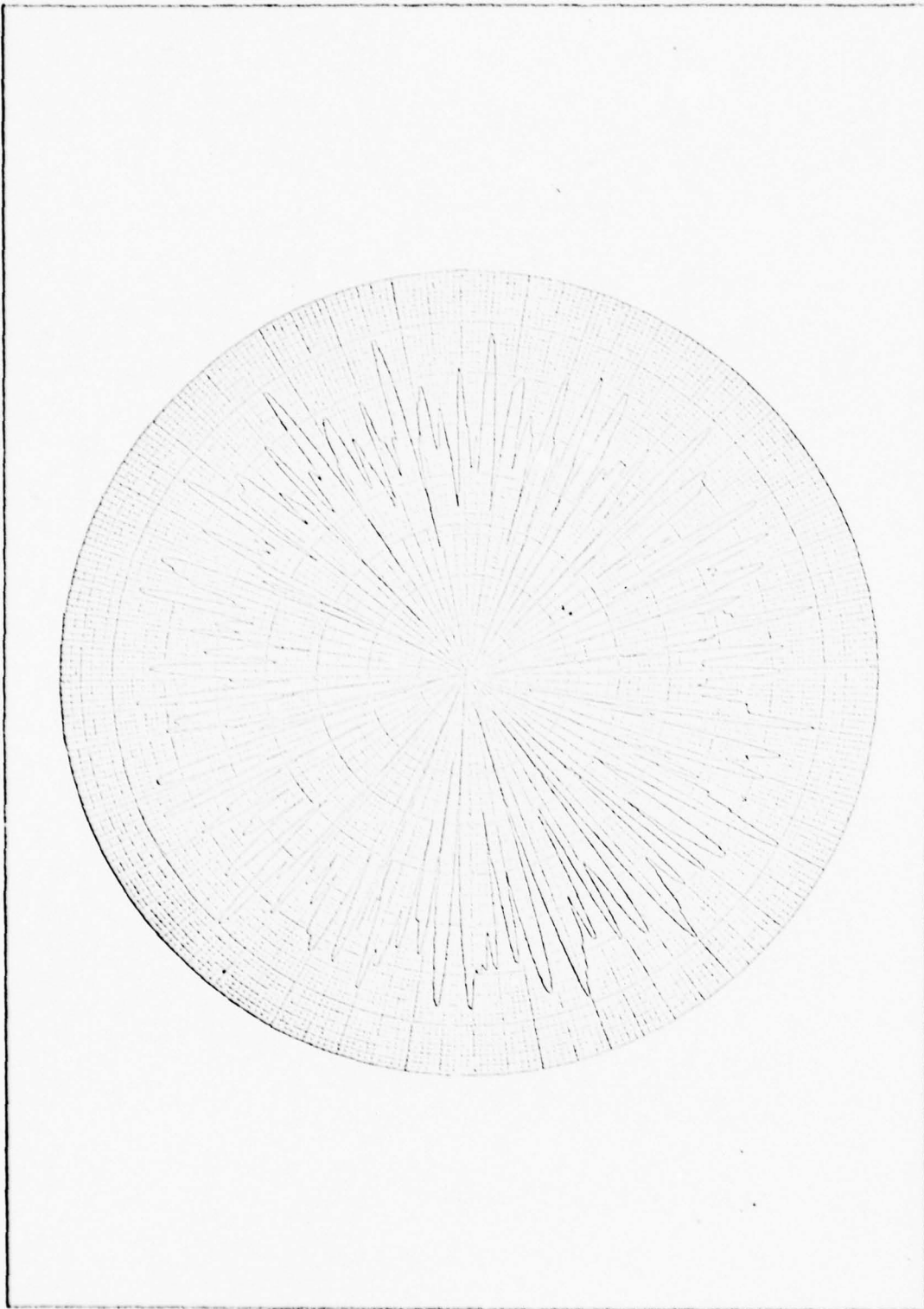
Fig 4.  Radar Cross Section of Tractor Tire vs Rotation Angle (Ref 20)

where $M(t)$ is a complex modulating function with period $T$. Because $M(t)$ is periodic, it can be represented as a Fourier series

$$M(t) = \sum_{i=-\infty}^{\infty} F_i \exp(j2\pi it/T) \tag{8}$$

Thus the backscattered field is of the form

$$E_s(t) = \sum_{i=-\infty}^{\infty} F_i \exp\, j(\omega_0 + 2\pi i/T)t \tag{9}$$

which has a power spectrum

$$S(\omega) = 2\pi \sum_{i=-\infty}^{\infty} |F_i|^2 \delta[(\omega - \omega_0) - 2\pi i/T] \tag{10}$$

where $\delta(\cdot)$ represents the Kronecker delta. Thus the power spectrum of the backscatter from a rotating polygonal cylinder is discrete, being nonzero only at frequencies given by

$$\omega = \omega_0 + iN\alpha\ , \quad i = 0, \pm 1, \pm 2, \ldots \tag{11}$$

Fig 5 depicts such a spectrum computed by Chuang using the Geometric Theory of Diffraction (Ref 71).

Four observations about this type of spectrum that are useful in a target recognition context are:

1. The spectrum is symmetric about the carrier frequency if the polygon is regular; however, it will not generally be symmetric for objects of arbitrary cross section.

2. The component at the carrier frequency tends to be the maximum signal, although it is not always.

3. As $|i|$ increases, the spectrum tends to decrease monotonically in magnitude.

4. If we consider the cylinder to have a translational velocity in addition to its radial velocity, then the amplitude spectrum drops sharply when

$$|i| > \frac{\omega_D}{N\alpha} \tag{12}$$

23

POWER SPECTRUM (dB)

0
-10
-20
-30
-40

$a = 10\lambda_0$
$N = 10$
TM1

$m = (\omega - \omega_0)/N\alpha$

-10  -8  -6  -4  -2  0  2  4  6  8  10
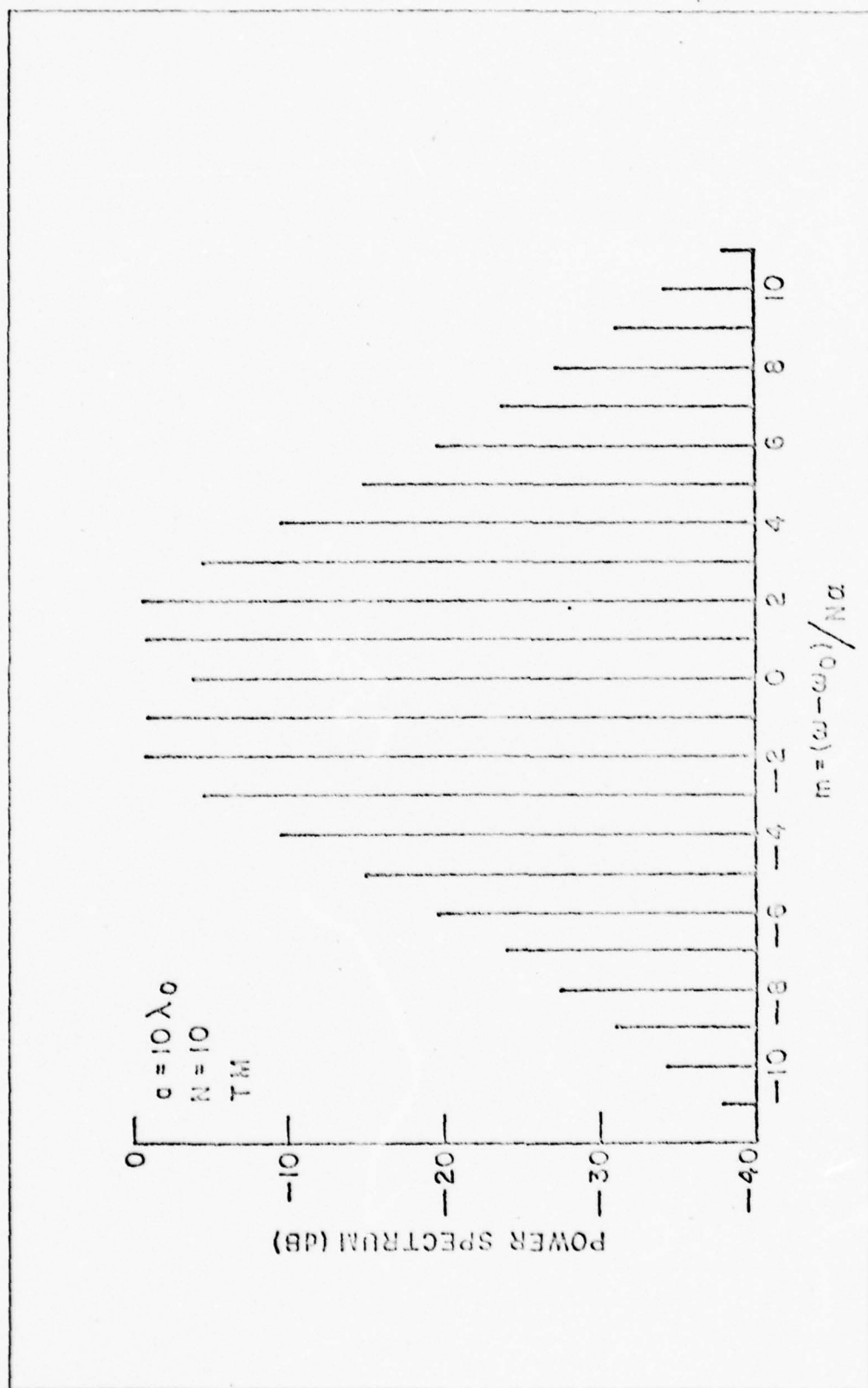
Fig 5.  Power Spectrum of a Rotating Polygon (Ref 11)

where $\omega_D$ is the doppler radian frequency. If a is the maximum radius of the rolling cylinder and c the speed of light, then

$$\omega_D = 2a\omega\omega_o/c$$

$$= \frac{4\pi v_R}{\lambda} \tag{13}$$

where $v_R$ is the target radial velocity.

To design an optimal classifier for simple PAM targets under the assumption of no unknown parameters would be rather easy. For the type of situation discussed, the amplitudes of the various spikes could be precomputed for frequencies out to a certain harmonic, and a simple correlation procedure could be implemented using the power spectral density of the received signal. If the received signal is first demodulated to baseband the power spectrum could be written as

$$S(\omega) = \sum_{i=-\infty}^{\infty} F_i \delta(\omega - \omega_D - i\Delta\omega) \tag{14}$$

where the pitch $\Delta\omega = 2\pi/T$, and the $F_i$ represent the amplitude of the spikes. For the present, receiver noise will be neglected. Using observation 4 above, $S(\omega)$ may be approximated by retaining the spectrum between 0 and $2\omega_D$:

$$S(\omega) = \sum_i F_i \delta(\omega - \omega_D - i\Delta\omega) \tag{15}$$

where i ranges over all integers such that

$$|i| < \frac{2\omega_D}{\Delta\omega} \tag{16}$$

There are numerous target parameters that will affect the power spectrum of the received signal, four of which are range, elevation angle, azimuth angle and velocity. Range changes alter the $F_i$, but, as long as the far field assumption may be maintained, the range variation simply acts as a multiplicative factor equally applied to all $F_i$. Thus range variation may be removed by simply normalizing by the peak amplitude or the square root of the signal energy.

25

For the simple type of scatterer considered thus far, the elevation angle between radar and target affects only $\omega_D$, since $\omega_D$ is proportional to $v_R$ which is the vector inner product of radar boresight and target velocity. For more complex targets, however, the $F_i$ are a function of elevation. Consider, for example, that the rotating structure represents the partially exposed running gear of a vehicle. At certain elevations more energy will be scattered by the rotating body than at others.

Both $F_i$ and $\omega_D$ are functions of the azimuth angle between the radar and target. The $F_i$ will vary with azimuth since the RCS of the scatterer will change, and $\omega_D$ will change since the radial velocity is a function of target aspect.

Finally, the whole spectrum will vary in an accordion-like fashion with the target velocity. If the elevation angle $\phi$ is defined as the angle measured from the horizon to the radar boresight, and the azimuth angle $\theta$ is defined as the angle measured clockwise from the target velocity vector to the radar boresight, the radial velocity may be written as:

$$v_R = v \sin \theta \sin \phi \tag{17}$$

Then the doppler radian frequency becomes

$$\omega_D = \frac{4\pi}{\lambda} v \sin \theta \sin \phi \tag{18}$$

and the pitch is

$$\Delta\omega = \frac{Nv}{a} \tag{19}$$

For fixed target parameters, letting r represent represent range, the power spectrum can be expressed as

$$S(\omega, r, \theta, \phi, v) = \sum_i F_i(r, \theta, \phi) \delta(\omega - \frac{4\pi v}{\lambda} \sin \theta \sin \phi - \frac{iNv}{a}) \tag{20}$$

where i ranges over all integers such that

$$|i| \leq \frac{8\pi a}{N\lambda} \sin \theta \sin \phi \tag{21}$$

The discussion to this point has assumed that all target parameters are constant. If any of the target parameters vary, as they invariably will for any real radar target, the received signal can no longer be assumed stationary, and the formal requirements for the existence of the power spectral density are no longer met. The short-time spectrum can, however, still be computed and will be a useful representation of the target signature if the observation interval is chosen sufficiently short to assume local stationarity.

Data Bases

Two separate data bases were available for analysis. Both consisted of the time domain signatures of vehicles driven through the radar beam transmitted from fixed antennas. The two sets are referred to as Data Bases A and B. Of the two, Data Base A was the most complete and appeared to be the most reliable.

Data Base A. The radar used to collect the data for Data Base A was a high pulse repetition frequency (PRF), coherent system. The frequency transmitted was sufficiently high to be in the optical region (where the scatterer's dimensions are large with respect to wavelength) of the targets used. A block diagram of the radar is shown in Fig 6.

The stable local oscillator (stalo), shown at the upper left side of the figure, provides the stable reference frequency used for both the transmitter and receiver. The transmitter signal is shifted upward in frequency by $f_i$, the intermediate frequency (IF) of the clock, in the transmitter mixer. This signal is filtered, amplified, and gated on and off at the PRF to form the transmitted signal which is radiated toward the target.

The received signal passes through the antenna to a circulator and then to radio frequency (rf) amplifiers and the first mixer. The
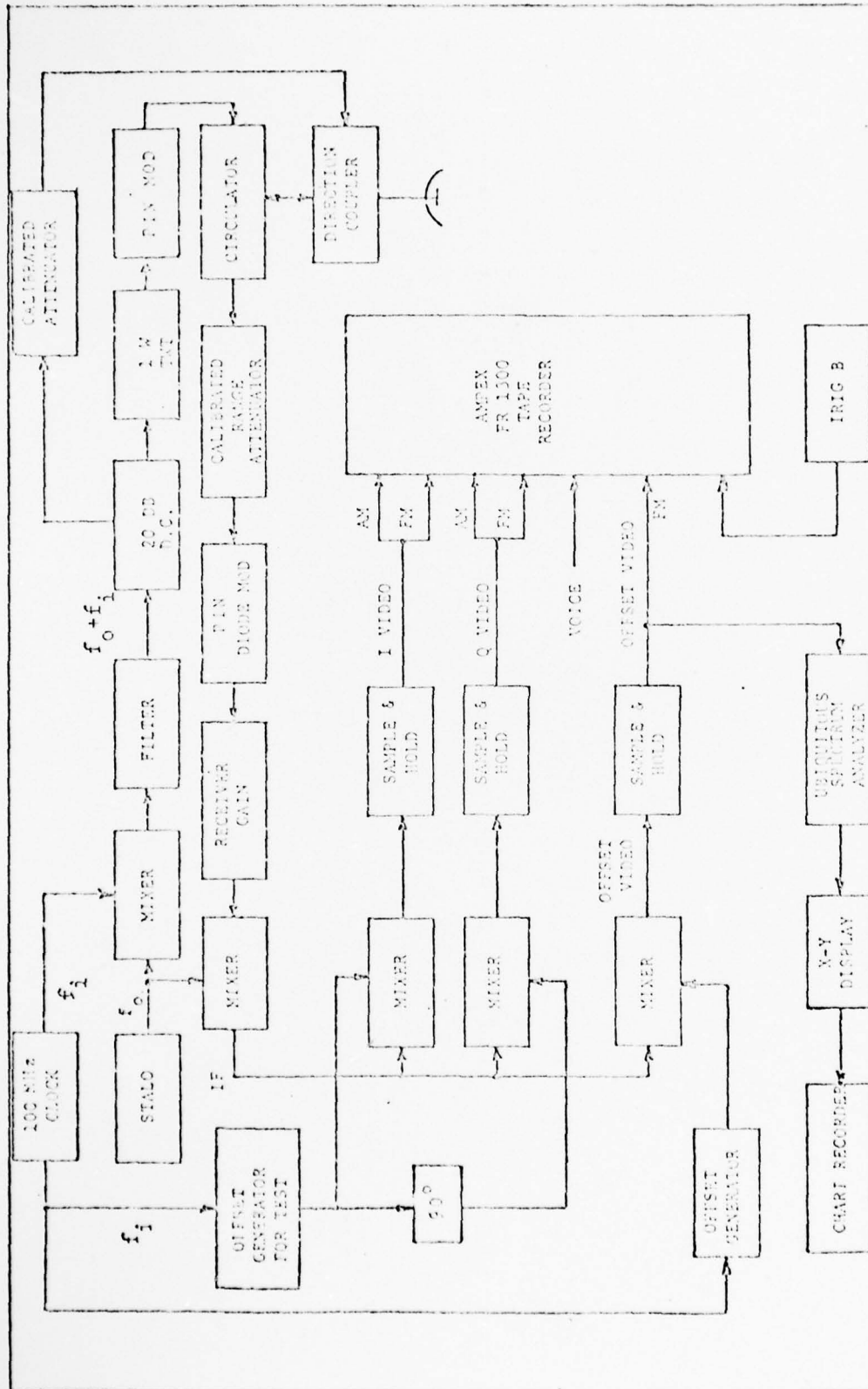
Fig 6.  Block Diagram of Data Base A Radar

reference frequency to the first mixer is the stalo output, $f_o$. The output of this mixer is the IF signal at $f_i$. The output signal of the mixer for the offset video channel is on a low frequency carrier. After amplification, the offset video signal is passed through boxcar circuits, the timing of which determine the range gate. Following more amplification, the offset video signal is recorded and its spectrum is displayed on the spectrum analyzer. The spectrum being shifted away from zero doppler permits its being observed without folding.

To preserve the phase information in the signal, in-phase (I) and quadrature (Q) video channels are required; hence, the IF signal passes through two additional mixers. The references for these mixers are at $f_i$ but are 90 degrees out of phase with one another. The outputs of these channels are amplified, passed through the boxcar circuits, and recorded. The I and Q video channels are matched in amplitude.

The tape recorder used was an analog Ampex FR-1300 which has seven data channels plus a voice channel. With the tape speed used, 60 inches per second, the recorded band is 0 to 20 kHz on the FM channels with a signal-to-noise ratio of 44 dB and a harmonic distortion of two percent or less. Although the FM channels were the primary data channels, the I and Q video were sometimes recorded on the greater bandwidth AM channels as well. This was done to insure that no high frequency information was lost.

Signatures were recorded of three separate target vehicles under a variety of conditions. The target vehicles will be referred to as Target 1, Target 2, and Target 3. Target 1 had running gear that was quite dissimilar to that of Targets 2 and 3, while the latter two were quite similar in that respect. Approximately 200 runs each were recorded for Targets 1 and 2, while 35 were made for Target 3.

Signatures were recorded for Targets 1 and 2 positioned on a rotary platform. The vehicles were placed on jacks so that the running gear could turn freely without the vehicle moving off of the platform. Antenna depression angles and target slant ranges used were $4.5°$ at $27.6$ m and $26.3°$ at $41.5$ m. The rotary platform data was not used in the pattern recognition analysis reported here. It was used to gain some insight into the nature of the signal modulation produced by the vehicles. This part of the data is free of one source of noise, specifically that caused by the vehicle traversing rough terrain.

The data that was used in the pattern recognition analysis was taken with each of the three targets moving through the radar beam in the center of a grassy field. This field data was taken with an antenna depression angle of $3.5°$ and a slant range of $325$ m, where slant range is defined as the distance from the radar antenna to the target.

Both rotary platform and field runs were taken at $22.5°$ degree azimuth angle intervals beginning with the vehicle headed toward the radar. The returns were monostatic. The antennas used were linearly polarized, and both horizontal and vertical samples were taken. The horizontally polarized returns were used in this analysis since no vertically polarized returns from Target 3 were recorded. It should be noted, however, that the level of sideband modulation compared to skin line amplitude appeared to be slightly higher for vertical polarization. During the field runs, the vehicles were sometimes at relatively constant velocity, and at other time they were accelerating.

Data Base B. The experimental set up for the second set of data was similar to that for Data Base A except that the radar used was CW instead of pulsed. Also, the runs were only taken at every $45°$ of

30

azimuth, and there were no platform runs. Since the radar was not coherent and no offset was used, there may have been some frequency foldover in the spectra, but it was neglible. Observation 4 of the previous section indicates that the target spectrum should drop sharply at zero Hz, and this was confirmed by examining the spectra of live data. Thus any foldover of significant amplitude will be caused by negative doppler clutter which will be concentrated at very low frequencies. This low frequency foldover is of no concern since the clutter so thoroughly dominates the signal at these frequencies that this band is useless anyway.

Data Base B consisted of returns from five different vehicles, Targets 1, 2, 4, 5, and 6. Targets 1 and 4 had similar running gear and produced comparable signal modulation, while Targets 2, 5, and 6 were from the same general category. The Targets 1 and 2 vehicles used in the Data Base B measurements were not the exact same ones used for Data Base A, but they were of the same respective model.

## III The Optimal Classifier

This chapter presents optimal classification techniques, including Bayes classification, composite hypothesis testing, and sequential classification. It is concluded that such techniques, although conceptually straightforward, are not directly applicable to the problem at hand because of the lack of a complete physical model and complexities in the data and required implementation.

### Bayes Classifier

Automatic target identification may be considered to be an application of statistical hypothesis testing. In such a context, it is well known that the Bayes classifier is optimal in the sense that it minimizes the expected risk or probability of error (Ref 25:89). This optimality is due to the fact that the Bayes classifier uses all of the statistical information from the problem as efficiently as possible. A Bayes formulation implies that the identification problem may be cast strictly in statistical terms and that all statistical information about the problem is known. Neither of these conditions may be strictly true, but as in any complex physical problem certain simplifying assumptions may be made that will allow us to proceed. For example, the class conditional densities are generally unknown but may be approximated by one of two different ways or by a combination of the two. One method is to attempt to deduce the relevant statistics from a careful study of the underlying physics of the process. The second method is to ignore the physics and simply use the statistics of the training set in an attempt to estimate the class conditional densities using some method such as Parzen windows (Ref 18:88-95).

In its most abstract setting any target identification scheme is a rule for assigning any observed target signature to a class:

$$A:B \rightarrow \Omega \tag{22}$$

32

where A is the identification rule, B is the observation space, and $\Omega=\{\omega_1,\omega_2,\dots,\omega_m\}$ is the finite set of target classes. (It is possible that one of the $\omega_i$ may represent a rejection class, i.e. no decision is made.) Thus, from a geometric point of view, the decision rule partitions the observation space and assigns class labels to the various regions:

$$A:B_i \rightarrow \omega_i, i=1,2,\dots,m \tag{23}$$

where

$$B_i \cap B_j = \emptyset, \forall\, i \neq j \tag{24}$$

and

$$B = \bigcup_{i=1}^{m} B_i \tag{25}$$

In the general Bayesian formulation, a cost is assigned to each type of decision:

$$c_{ij} = \text{cost of deciding } \omega_i \text{ when } \omega_j \text{ is true} \tag{26}$$

If a symmetric cost assignment is made, i.e.

$$c_{ij} = \begin{cases} 1 & , \quad i \neq j \\ 0 & , \quad i = j \end{cases} \tag{27}$$

then the Bayes decision rule is to select $\omega_j$ for which

$$P(\omega_j|X) > P(\omega_i|X) \quad , \forall\, i \neq j \tag{28}$$

where X represents an observation. In this discussion $P(\cdot)$ is used to designate a probability and $p(\cdot)$ represents a probability density function. In words, the rule simply says to take an observation and assign it to the class for which the posterior probability is the greatest; therefore, this special case is frequently called the maximum a posteriori (MAP) classifier. By applying Bayes' theorem, the posterior probabilities may be calculated from the prior probabilities $P(\omega_i)$ and the class conditional densities $p(X|\omega_i)$:

$$P(\omega_i|X) = \frac{P(\omega_i)p(X|\omega_i)}{p(X)} \tag{29}$$

where

$$p(X) = \sum_{i=1}^{m} P(\omega_i) p(X|\omega_i) \qquad (30)$$

The decision rule then becomes to choose $\omega_j$ such that

$$P(\omega_j) p(X|\omega_j) > P(\omega_i) p(X|\omega_i) , \forall i \neq j \qquad (31)$$

For this multihypothesis problem, when the prior probabilities are equal, it is frequently convenient to use the likelihood ratio test: choose $\omega_j$ such that

$$\ell(X) = \frac{p(X|\omega_j)}{p(X|\omega_i)} > 1, \forall i \neq j \qquad (32)$$

If in addition to equal prior probabilities, the class conditional densities are of an exponential form, then the log likelihood ratio test is normally used: choose $\omega_j$ such that

$$\ln \ell(X) = \ln p(X|\omega_j) - \ln p(X|\omega_i) > 0, \forall i \neq j \qquad (33)$$

For general cost functions, the expected value of the cost or Bayes risk may be written (Ref 69:47) as

$$R = \sum_{i=1}^{m} \sum_{j=1}^{m} c_{ij} P(\omega_j) P(\text{deciding } \omega_i | \omega_j \text{ true})$$

$$= \sum_{i=1}^{m} \sum_{j=1}^{m} c_{ij} P(\omega_j) \int_{B_i} p(X|\omega_j) \, dX \qquad (34)$$

It is well known that the Bayes risk for this general case may be minimized by selecting the $B_i$ such that

$$\sum_{j=1}^{m} c_{ij} P(\omega_j) p(X|\omega_j) < \sum_{j=1}^{m} c_{jk} P(\omega_j) p(X|\omega_j), \forall k \neq i \qquad (35)$$

see, for example, Fukunaga (Ref 25:74-75) or Van Trees (Ref 69:46-52).

If the PAM targets were always to be identified with fixed, given parameters, and it could be assumed that the only source of uncertainty in the signatures was additive white Gaussian noise, the form of the

31

Bayes classifier would be easy to specify. Under these conditions, it is well known (Ref 25:90-91) that the Bayes classifier may be implemented using matched filters. Such a classifier consists of a bank of matched filters, one for each type of target, followed by a comparator network to choose the largest output. The target would be assigned to the class whose matched filter gave the largest output.

If sources of uncertainty that are present in the actual data are added, this simple matched filter realization is no longer optimal. As a real target traverses the terrain, irregularities in the surface cause changes in the target parameters resulting in modulation in both amplitude and position of the sideband spikes. Furthermore, if the classification is done digitally, distortions due to discretization and finite observation periods occur. Such digitally induced distortion will be discussed in greater detail in the following chapter. Finally, it is necessary to be able to perform the identification process with generally varying, unknown target parameters.

## Composite Hypothesis Testing

As was noted in Chapter II, the signature of a PAM target depends not only upon the type of target but also upon the target parameters, range, velocity, and aspect angles. These target parameters are transformed by the electromagnetic scattering to signal parameters. The signal parameters, conditioned on the type of target being illuminated, specify the magnitudes and locations of the discrete Fourier components of the spectral signature of a simple PAM target. For this simplified example, the received signal for each class would be known except for a finite set of parameters and noise. This is the composite hypothesis testing formulation (Ref 69:86-96). In the following discussion the set of unknown signal parameters are represented by the vector $\mu$.

35

If all of the probability densities can be specified, the problem may be reduced to a simple hypothesis testing one by integrating over the parameter space

$$p(X|\omega_i) = \int_\chi p(X|u,\omega_i)p(u|\omega_i)d\,u \tag{36}$$

where $\chi$ represents the parameter space. Then Equation (31) may be used to make an optimal decision. When it is desired to estimate $\mu$, this problem is referred to as simultaneous detection and estimation in the communications theory literature. Entire dissertations have been devoted to the solution of this problem, e.g. Gobien (Ref 28).

Under certain circumstances, it may be assumed that the radar processor maintains a track file on the target, and good estimates of all the parameters are available. The decision rule for the multiclass problem with a symmetric cost function then becomes to choose the $\omega_j$ for which

$$P(\omega_j)p(X|\omega_j,\mu) > P(\omega_i)p(X|\omega_i,\mu) \quad ,\forall\, i \neq j \tag{37}$$

The additional knowledge about the random parameters is simply incorporated in the class conditional densities.

Robinson and DeNuzzo (Ref 57) argue persuasively that it may not be possible to evaluate class conditional densities as in Eqs (36) or (37). They suggest instead that the prior probabilities be modified. The prior probability would be multiplied by a fuzzy set membership function (Ref 71) for each known parameter. For example, if the target velocity is known to be in the middle of the operating range for class 1 and toward the high end for class 2, the fuzzy set membership function for that velocity might be unity for class 1 and 0.5 for class 2.

Another practical alternative would be to normalize the signature with respect to certain parameters. For example, target range may not provide any useful discriminating information. As long as the target

36

is in the far field, range only affects the signal amplitude and, consequently the signal to noise ratio. Then the signal could be energy normalized without losing any information, assuming the signal to noise ratio is acceptable, and target range would be removed as a parameter.

Another practical approach would be to increase the number of classes. A signature from Target 1 at zero degree azimuth may not resemble a signature from Target 1 at 135 degrees azimuth. Thus Target 1 may be decomposed into several subclasses depending on the known azimuth angle. The most useful approach is to consider the effect of each parameter separately and apply whichever technique is most appropriate.

If the problem were to identify long, rolling polygons as described in Chapter II, a theoretically optimal classifier could be designed. Unfortunately the rolling polygonal scatterer only qualitatively simulates the running gear of an actual vehicle. Several sources of uncertainty arise in the actual data that do not exist in the simplified model. At certain aspects the rotating structures will cause specular flashes. Rapid changes in pitch, yaw, roll, and linear velocity due to interactions between the vehicle, terrain, and the human operator cause complicated modulations of the signal. Also, random modulation due to vibrations is unaccounted for in the PAM model. Since the state of the art of numerical methods as applied to electromagnetic scattering problems is not sufficiently advanced to account for such complexities, no complete, quantitative model will be forthcoming in the near future.

The only practicable method of estimating the class conditional probability densities that are required for optimal classification would be through an extensive measurement program. The available data bases are not extensive enough to provide good estimates, and a sufficiently

comprehensive measurement program could be exorbitantly expensive.
Because of these practical difficulties, the approach taken in the sub-
sequent chapter is to extract features that are relatively invariant
with respect to target parameters.

## Sequential Classification

When attempting to identify a moving radar target, it is possible
to take multiple observations to increase the reliability of the
estimate of the target class. The Wald sequential probability ratio
test (Ref 25:77-84) may be used to reduce the probability of error to
zero if enough observations are available. The Wald test has only been
applied to a two-class problem and, as with all optimal procedures,
assumes all of the underlying statistics are known.

The sequential observation classification technique adopted in the
sequel, is a simple plurality-vote scheme. Since no effort was made to
estimate the reliability of the decision made after a number of obser-
vations were taken, the technique might be more aptly termed a
multiple observation classification scheme. The number of observations
assigned to each class is remembered, and the decision is made to assign
the target to the class with the largest number of votes. Shrihari
(Ref 61:151-179) has considered the theoretical properties of such a
voter in the context of a radar identification of aircraft problem.
It must be assumed that the radar is maintaining track on a given target,
in order to make the voter decision valid.

## IV  A Suboptimal Frequency Domain Classifier

This chapter presents the design of a suboptimal classification scheme. The classifier is suboptimal in the sense that the error rates achieved may be greater than the irreducible error rate. The short-time spectra of the target signatures are computed via the FFT. Features are then extracted that are relatively invariant to target parameters. Target identification is performed using linear discriminant analysis and nearest neighbor classification on the extracted features.

### Preprocessing

In the pattern recognition problem, preprocessing of the data enjoys an eminent position. Since the preprocessing is the first operation performed on the sensed data, and since it includes potentially noninvertible transformations (specifically ones that are not mono-morphic) any losses of information or distortions will be propagated through the entire classification process. Preprocessing may be formally defined as the transformation from the measurement space to the pattern space.

The input to the preprocessor is typically a noisy analog signal, the output of the sensor. The totality of all possible output signals from the sensor, the measurement space, may be characterized as a finite power space or a finite energy space, as appropriate. Since the subsequent operations are usually done digitally, the first preprocessing transformations usually consist of prefiltering and sampling the time domain data. The prefiltering is performed, of course, to prevent aliasing and to reduce out of band noise, and the Nyquist criterion must be observed when selecting the sampling rate.

The data used in this experiment were all recorded on analog tape using frequency modulation with a 20 KHz bandwidth. To determine an

appropriate sampling rate, selected data runs were digitized at 40 KHz and transformed using an FFT. The resulting frequency domain signals were plotted and examined. By visual inspection, it was determined that one KHz of bandwidth would be sufficient to capture the information of interest.

The data was put through an analog, two pole filter which is down about 3 dB at one KHz. The antialiasing-filter frequency response is depicted in Fig 7. The prefiltered signal was sampled at two KHz. On data base A, where both the in-phase and the quadrature signals were recorded, both channels were sampled at two KHz resulting in an unfolded spectrum two KHz wide centered at zero Hz. For the actual implementation of a classifier of the type being discussed, in order to minimize computational requirements, a phase locked loop could be used to track the skin line (dominant pole) of the signal. The sampling rate could be set to four times the skin line frequency.

If the features are to be extracted from the time domain signal, no further preprocessing may be required. It is frequently true, however, that the sampled time domain is not an appropriate pattern space. Then some further preprocessing transformation is required. As Pavlidis (Ref 53:2-3) has observed, if the information pertinent to the classification task tends to be time-limited, for example isolated time domain pulses such as electrocardiograph data, then the time domain is an appropriate pattern space. If, on the other hand, the essential information tends to be spread out in the time domain, it will be band-limited in the frequency domain. Thus the discrete Fourier transform (DFT) may be a candidate for the final preprocessing transformation.

Other orthogonal transforms that have found use in the pattern recognition context are the Karhunen-Loève, the Walsh/Hadamard, and the
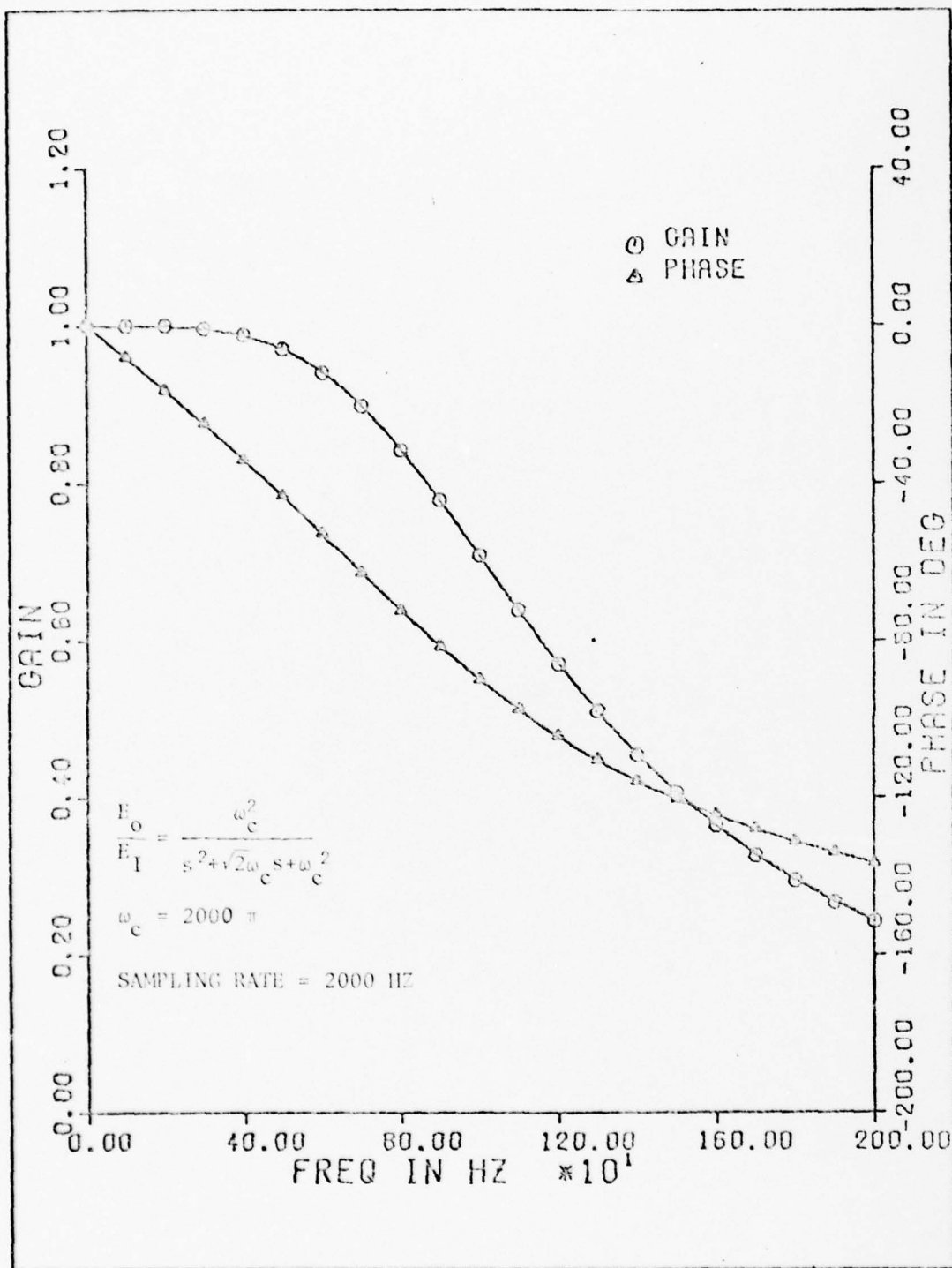
Fig 7. Prefilter Frequency Response

$$\frac{E_o}{E_I} = \frac{\omega_c^2}{s^2 + \sqrt{2}\omega_c s + \omega_c^2}$$

$\omega_c = 2000\pi$

SAMPLING RATE = 2000 HZ

41

Haar transforms. Andrews (Ref 1:212-215) provides a good comparison of the various transforms. Kabrisky and Carl have used a family of linear transforms in biologically motivated optical recognition schemes (Ref 37).

For the present application, the space of short-time Fourier spectra was chosen as the pattern space for several reasons. First, and most importantly, the Fourier spectrum of a moving, periodically amplitude modulated target is interpretable in terms of the physics of the target as was discussed in Chapter II. Second, with the advent of the FFT (Ref 12), the discrete Fourier spectrum can be computed efficiently with $N \log_2 N$ computer operations where N is the order of the FFT. Finally, it can be shown that the Fourier transform coverges nearly as rapidly as does the Karhunen-Loeve transform (Ref 31) which is an optimum representation in a minimum mean squared error sense.

If the usual communication-theory type assumption of stationary, white, Gaussian random processes could be made, conventional techniques (Refs 4 and 5:532-571) for power spectral density estimation could be applied. Then a simple matched filter type of classifier could be designed in the frequency domain. Since the radar returns from moving, amplitude modulated targets are certainly not stationary, such techniques are not directly applicable. In the previous chapter, a Bayesian classifier was discussed, under the assumption that if certain parameters were fixed, the time series was stationary. For now, however, the point of view assumed is that each contiguous short-time spectrum is simply an independent observation of the target.

The duration of the time record to be transformed via the FFT is of critical importance, since its reciprocal is the lower bound on the

achievable frequency resolution. Thus, an increase in the length of the time record considered results in finer frequency resolution; however, the longer the time record is, the more apt are the nonstationarities in the data to cause smearing in the short-time spectrum. Another obvious drawback to processing long time records is the increase in memory and computational power required. Since a resolution of about two Hz was deemed desirable, a time record of one half second was initially used.

An acceleration in the target's radial velocity during the integration time of the FFT results in a distortion of the spectral lines that characterize the target. Nichol (Ref 43) has described this effect in the analysis of spectrograms of the acoustic emissions of rotating machinery. As described in Chapter II, the time domain signature of an amplitude modulated, moving target tends to be of the form:

$$s(t) = \sum_{i=1}^{k} F_i(t) \exp j\omega_i(t)t + n(t) \tag{38}$$

where the frequencies involved are explicit functions of time to account for the effect of the target accelerations. Adopting the terminology commonly used in speech analysis, each "smeared" sinusoid in Eq (38) represents a formant. To simplify the following discussion, consider the normalized, noiseless, instantaneously monochromatic signal:

$$s_n(t) = \exp j\omega_n(t)t \tag{39}$$

Eq (39), representing a single formant, results from discarding the noise and other frequency components in Eq (38) and normalizing the sinusoidal amplitude to unity, tacitly assuming that this amplitude remains constant on the interval $[0,T]$, the FFT integration period. For sufficiently small T, target acceleration may be considered to be linear

and thus

$$\omega_n(t) = \omega_f + \gamma t \tag{40}$$

where $\omega_f$ represents the initial formant frequency and $\gamma T$ represents the total shift in frequency during the acquisition period.

The continuous, short-time spectrum of Eq (39) at $t = T/2$ is

$$s_n(\omega) = \int_0^T \exp j \, (\omega_f t + \gamma t^2) \exp(-j\omega t) \, dt \tag{41}$$

By completing the square in the exponent and changing the variable of integration, Eq (41) becomes

$$s_n(\omega) = \sqrt{\frac{\pi}{2d}} \exp(-j \gamma d^2) \int_{A-B}^{A+B} \exp(j \frac{\pi}{2} z^2) \, dz \tag{42}$$

where

$$d = \frac{\omega_f - \omega}{2\gamma} \tag{43a}$$

$$z = \sqrt{\frac{2\gamma}{\pi}} \, (d+t) \tag{43b}$$

$$A = \sqrt{\frac{2\gamma}{\pi}} \, (2d+T) \tag{43c}$$

$$B = \sqrt{\frac{\gamma}{2\pi}} \, T \tag{43d}$$

The integral in Eq (42) is the well known Fresnel integral that must be evaluated numerically or graphically using the Cornu spiral. Gersch and Kennedy (Ref 27) have evaluated this integral for various frequency shifts to describe the spectrum of sliding tones.

Fig 8 illustrates the distortion that results to the formants for six different shifts. This result, for both the continuous and discrete Fourier transforms of harmonically related formants, is due to Nichol (Ref 48). In this figure, the magnitude of the shift is proportional to the initial frequency of the formant and increases with the order of the harmonic. It can be seen that the height of the formant decreases with
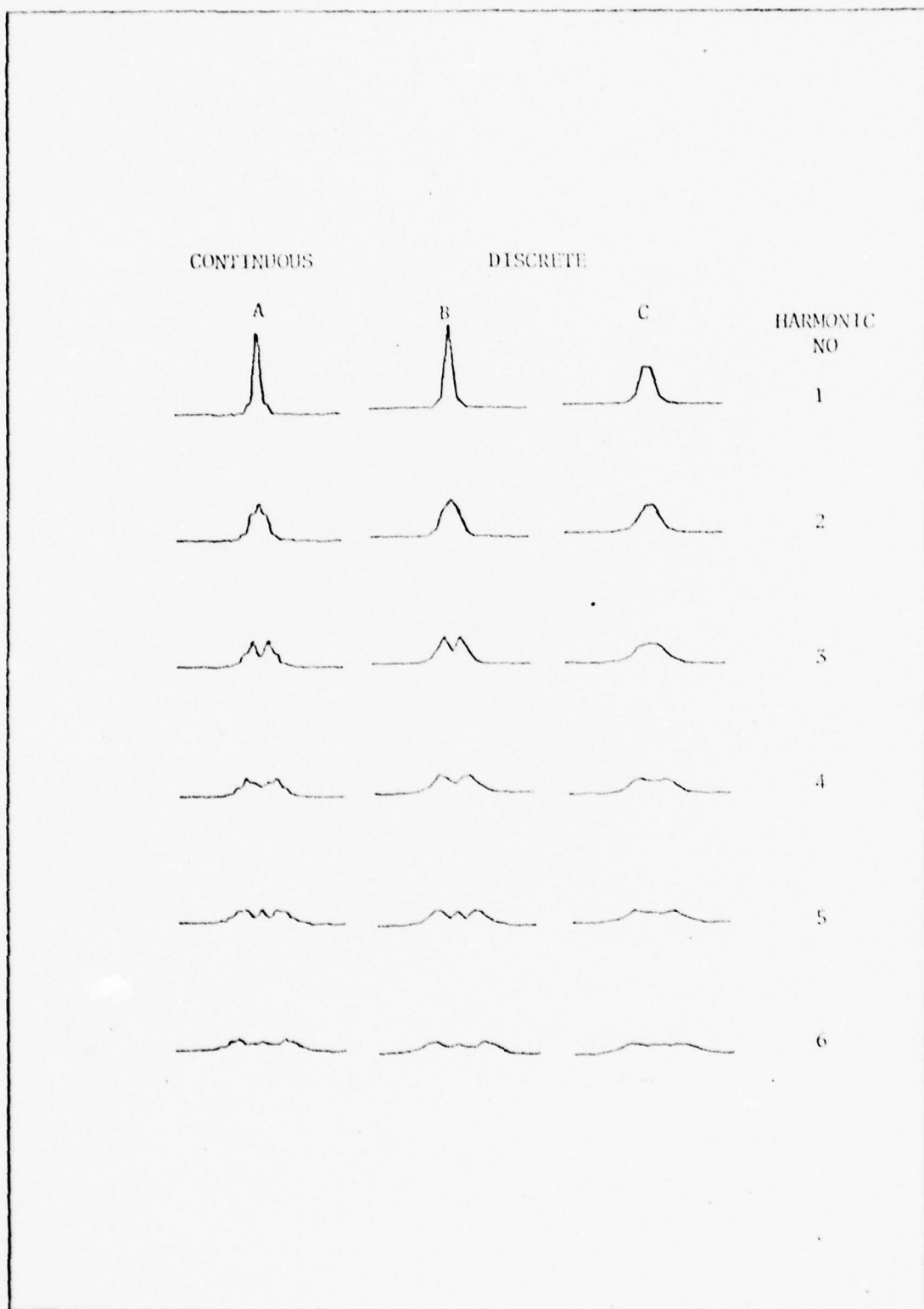
44

CONTINUOUS       DISCRETE

Fig 8.   Power Spectra of Nonstationary Harmonic Family (Ref 48)

increasing initial frequency, while the width increases, hence conserving energy. Columns B and C illustrate the result for the DFT, where the output is a function of the relationship between $\omega_f$ and T. Column B represents the result when

$$\omega_f = \frac{n2\pi}{T} \tag{44}$$

for n an integer. Column C illustrates the output when

$$\omega_f = (n+1/2) \, \frac{2\pi}{T} \tag{45}$$

The frequency spreading of the formants described above is frequently observed in the spectral plots of the experimental data. Fig 9 depicts the short-time spectrum of an accelerating target. Notice that the skin line at about 200 Hz exhibits the dual-peaking phenomena shown in Fig 8. Also, the two lower energy formants at about 375 and 400 Hz show the same type of frequency splitting.

The above discussion should illustrate the type of problem that is encountered in using short-time spectra of nonstationary time series for target identification. It becomes apparent that one should use as short a time record as possible to minimize distortions due to nonstationarities. If one were to consider the selection of the ideal time window length as an optimization problem, the cost function to be minimized could be written as

$$J(T) = G(T) + H(T) + I(T) \tag{46}$$

where the term G(T) is a positive, increasing function of T that represents costs due to nonstationary distortions. The function H(T) is also positive and increasing and typifies the computational costs involved. The final term I(T) is a positive decreasing function of T representing the frequency resolution costs. Because of the empirical nature and interdependence of the various parameters of Eq (46), the explicit forms
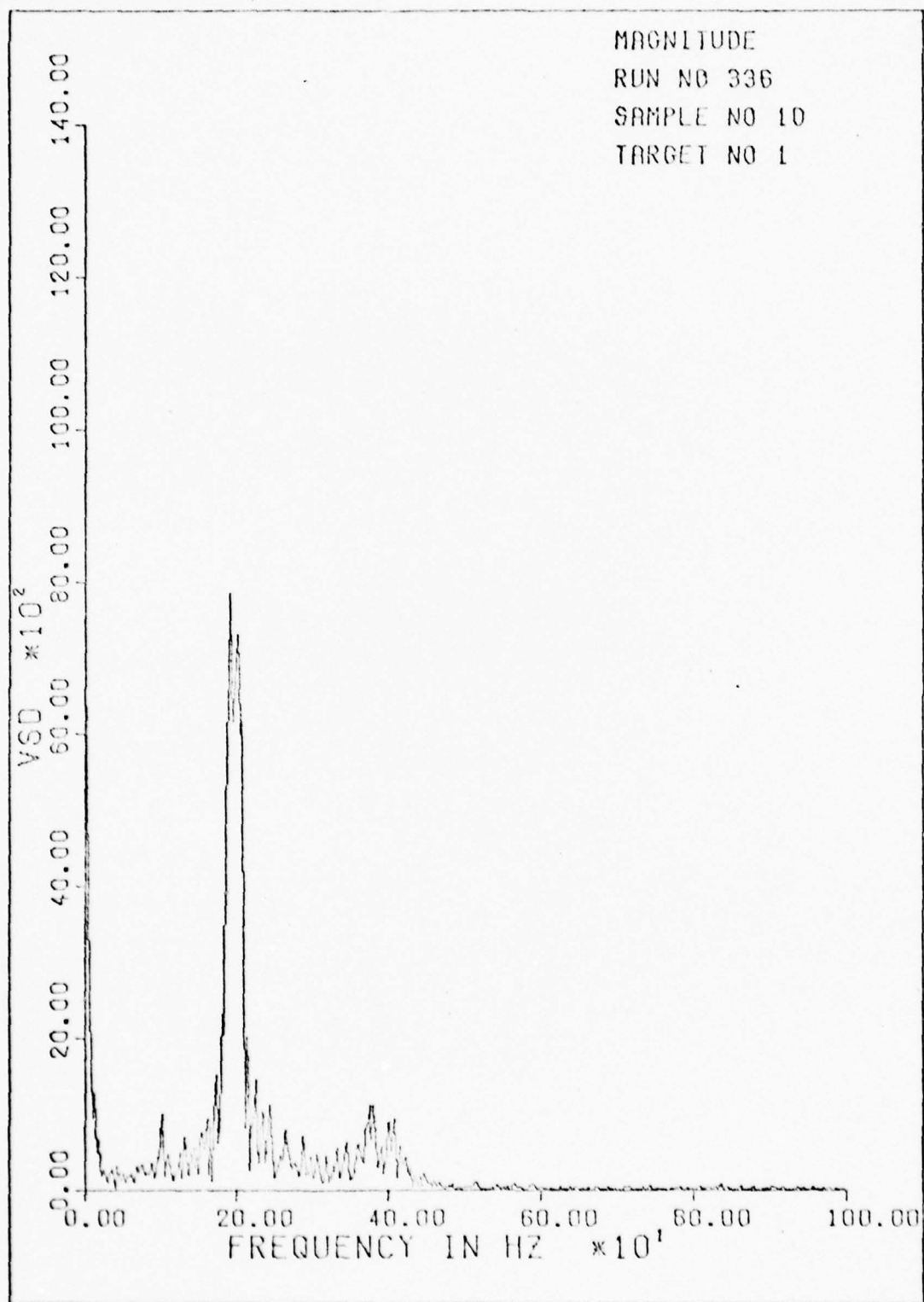
Fig 9.   Spectrum of an Accelerating Target

47

of the functions involved would be arbitrary and relating them to performance could be controversial, although the optimization theoretically could easily be performed once the functions were formally defined. Therefore the selection of T requires a certain amount of engineering judgement and a "cut and try" attitude.

Windowing. It is well known that the Fourier transform of the product of two functions is the convolution of the two individual Fourier transforms:

$$F[s(n)w(n)] = F[s(n)] * F[w(n)] \tag{47}$$

where the signals, convolution, and Fourier transform may be either continuous or discrete. This theorem has an important application in the computation of short-time spectra of amplitude modulated radar targets. Since only a finite record of the signal of interest will be transformed, the signal $s(n)$ is in effect being multiplied by a weighting function $w(n)$ which is nonzero on a finite interval.

If one simply truncates the time series at the end of the time record, by default, a rectangular weighting function has been applied:

$$w(n) = \begin{cases} 1 & 0 \le n < N-1 \\ 0 & n < 0, n > N-1 \end{cases} \tag{48}$$

where n is the index, and the time record consists of N points. The rectangular weighting function has the advantage of cheapness of implementation, and it has the narrowest attainable main lobe as illustrated in Fig 11. The achievable resolution of the FFT is determined by the width of the main lobe. On the debit side, the rectangular window has relatively high sidelobes, the first sidelobe being only 13 dB below the mainlobe in the power spectrum. As is apparent from Eq (47) and Fig 11, the convolution of the window spectral sidelobes with any spectral peaks from the signal will result in ringing or Gibbs phenomena.

This spectral distortion may be reduced by introducing a window

48

function that is smoothly tapered to zero. The tapering causes an increase in the width of the spectral main lobe and a resulting loss in resolution, but, at the same time, it increases frequency selectivity, i.e. the ability to resolve simultaneously signals of different amplitudes which are separated in frequency. Some commonly used windows are given below (Ref 4) and are shown in Fig 10, with the power spectra of each shown in Fig 11:

$$\text{Bartlett:} \quad w(n) = \begin{cases} \dfrac{2n}{N-1}, & 0 \le n \le \dfrac{N-1}{2} \\ 2 - \dfrac{2n}{N-1}, & \dfrac{N-1}{2} \le n \le N-1 \end{cases} \quad (49a)$$

$$\text{Hanning:} \quad w(n) = \frac{1}{2}\left[1 - \cos\left(\frac{2\pi n}{N-1}\right)\right], \quad 0 \le n \le N-1 \quad (49b)$$

$$\text{Hamming:} \quad w(n) = 0.54 - 0.46\cos\left(\frac{2\pi n}{N-1}\right), \quad 0 \le n \le N-1 \quad (49c)$$

Blackman:

$$w(n) = 0.42 - 0.5\cos\left(\frac{2\pi n}{N-1}\right) + 0.08\cos\left(\frac{4\pi n}{N-1}\right), \quad 0 \le n \le N-1 \quad (49d)$$

The Kaiser window (Ref 38) as defined by

$$w(n) = \frac{I_0\left[a\sqrt{m^2 - (n-m)^2}\right]}{I_0[am]}, \quad 0 \le n \le N-1 \quad (50)$$

has been shown to be optimum in the sense of maximum main lobe energy for a given peak side lobe amplitude. In Eq (50), $I_0(\cdot)$ is the modified zero order Bessel function of the first kind and $m$ equals $(N-1)/2$. The parameter $a$ may be adjusted to trade off resolution for selectivity. The obvious disadvantage of the Kaiser window is its computational complexity. For an exhaustive treatment of the various windows that have been used, the reader is referred to the recent paper by Harris (Ref 32).

Selection of an optimum window for a given application would entail minimizing a cost functional of the form
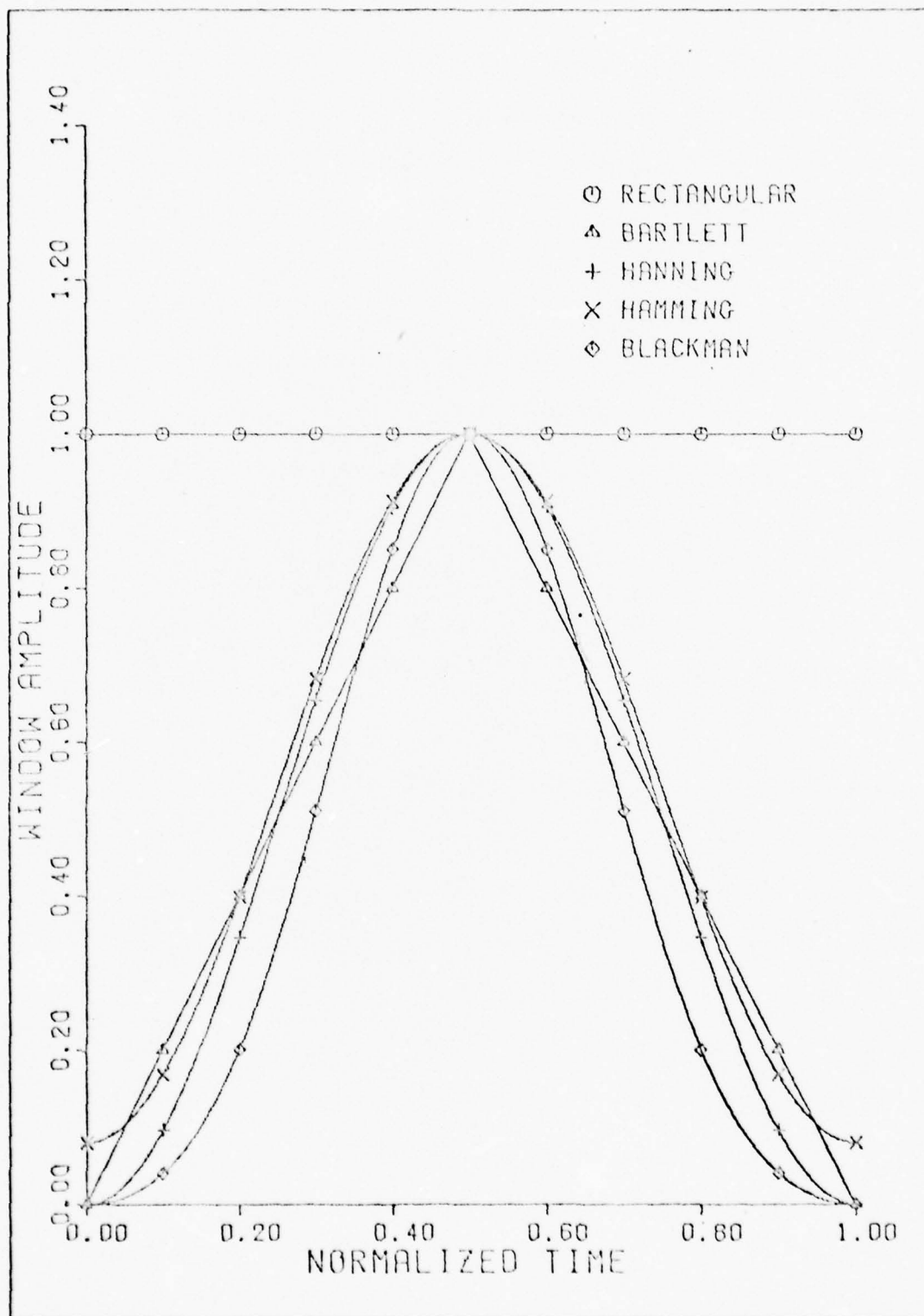
$$J(w) = G(w) + H(w) + I(w) \quad (51)$$

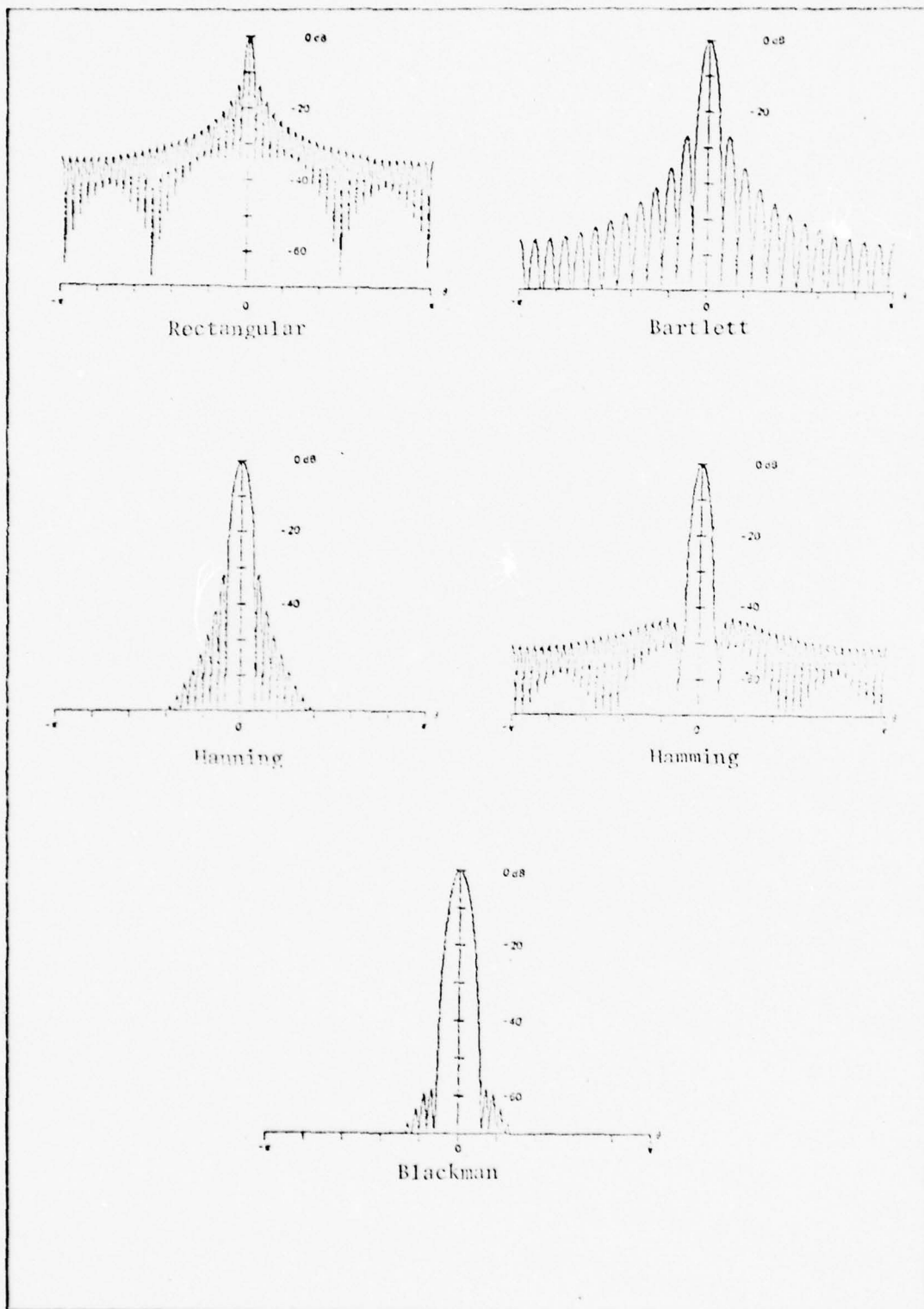Fig 10. Five Commonly Used Windows

Fig 11. Log Power Spectra of Five Commonly Used Windows (Ref 32)

where $G(w)$ is the resolution cost, $H(w)$ is the selectivity cost, and $I(w)$ is the computational cost. Rather than attempt to quantitatively define these costs and solve the optimization problem, an empirical approach was taken. Since the baseband representation of an amplitude modulated target was know a priori to be of the form

$$s(t) = \sum_{i=1}^{K} a_i \exp (j\omega_i t) \qquad (52)$$

it was possible to multiply this signal by various windows and numerically compute the resulting power spectra. The idealized log power spectrum

$$20\log_{10}|S(\omega)| = 10\log_{10}\left|\sum_{i=1}^{K} (2\pi a_i)^2 \delta(\omega - \omega_i)\right| \qquad (53)$$

is depicted in Fig 12.

Fig 13 shows the results of plotting

$$20\log_{10}|S(\omega)| = 20\log_{10}\left|F\{[s(n)+10^{-6}]w(n)\}\right| \qquad (54)$$

for each of the commonly used windows. The constant $10^{-6}$ was added to each term of the time series to scale the plot, thus -120dB corresponds to zero. It is readily apparent that the rectangular, Bartlett, and Hamming windows do not allow the spectrum to go to zero as rapidly as desired. Furthermore, the Bartlett window exhibits high quefrency ringing. (Queqrency is a measure of time associated with the Fourier transform of the log power spectrum of a signal). Since the ringing due to the sidelobes of the other windows is at too high a quefrency for the sampling rate, the Gibbs phenomena is not apparent in the other plots. The Hanning window was selected since its performance is satisfactory, and it is less than half as expensive to compute as the Blackman window. Since this simple window performed so well, no experiments were attempted with the more complex Kaiser window family.

using frequency modulation with a 20 kHz bandwidth.   To determine ...

39

IDEAL POWER SPECTRUM
TARGET NO 1
RADIAL VELOCITY 10 MPH

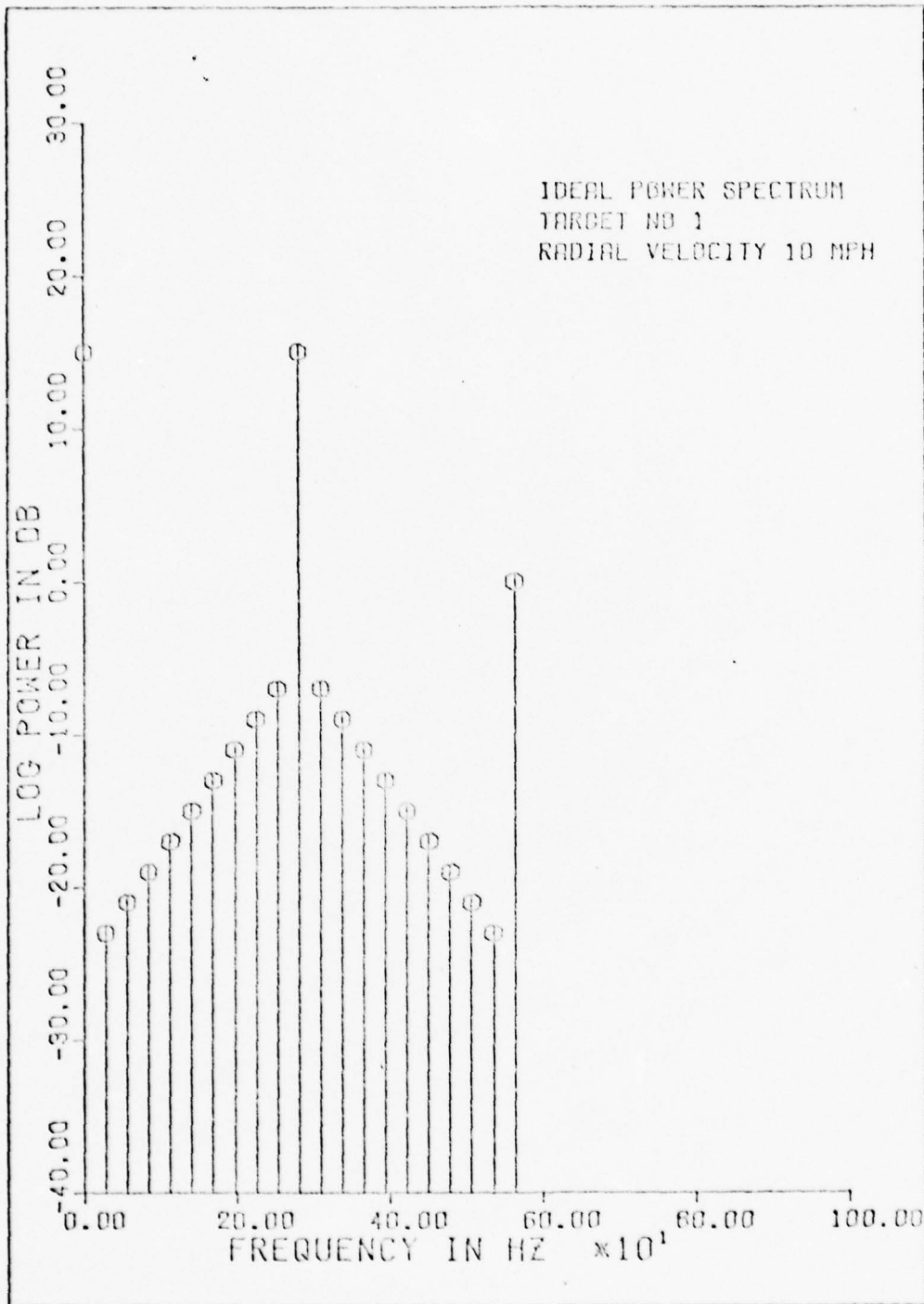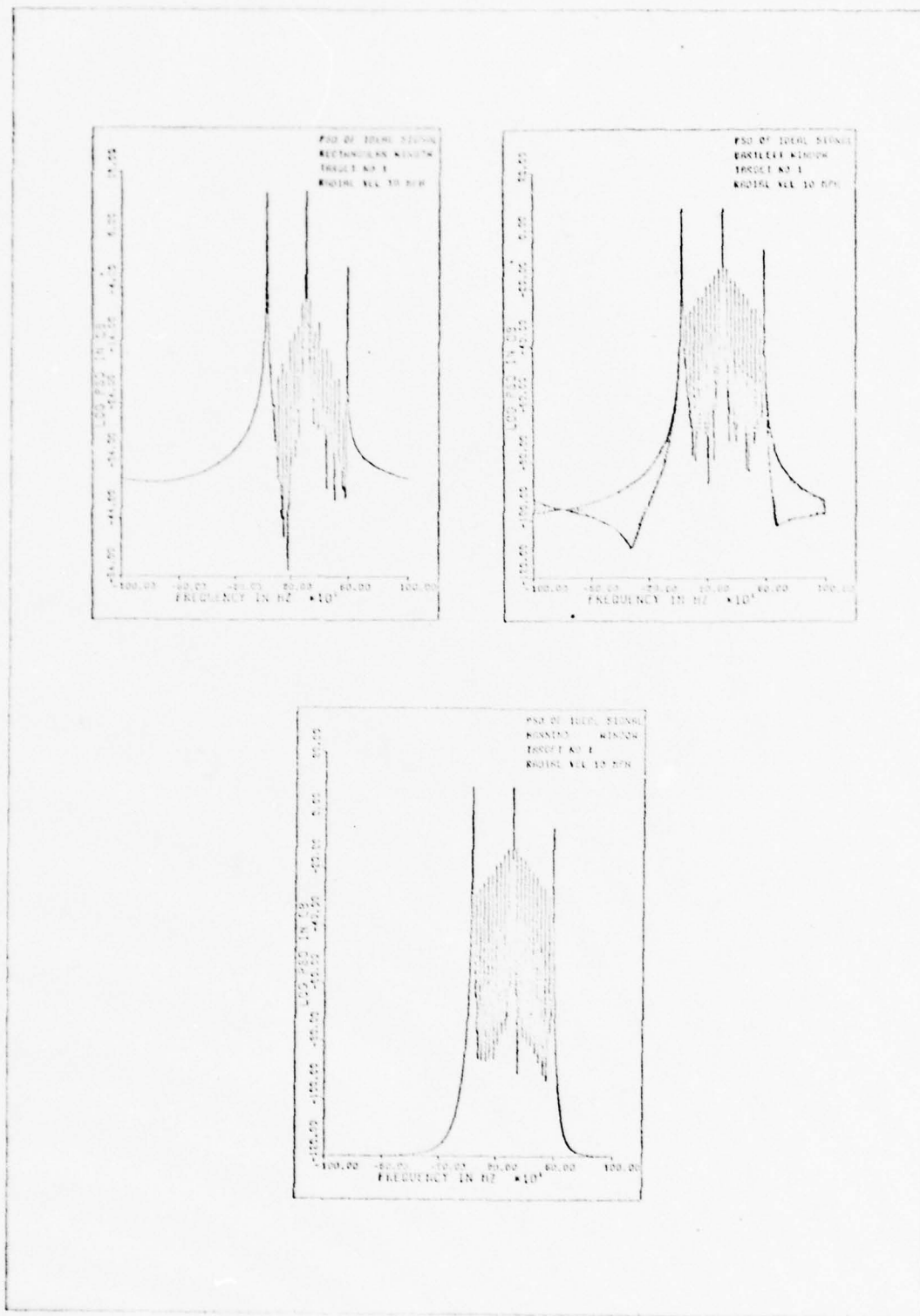Fig 12.  Idealized Target Spectrum

Fig 13.  Idealized Target Power Spectra Computed Using Various Windows
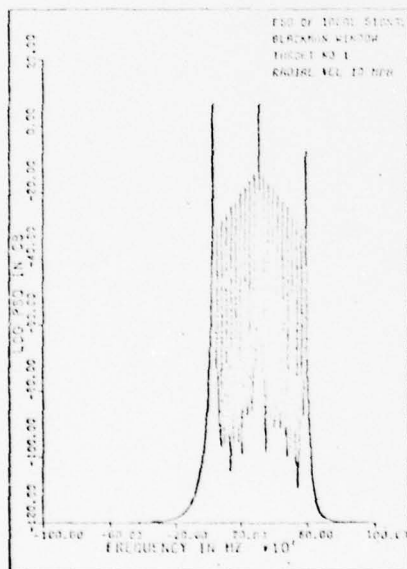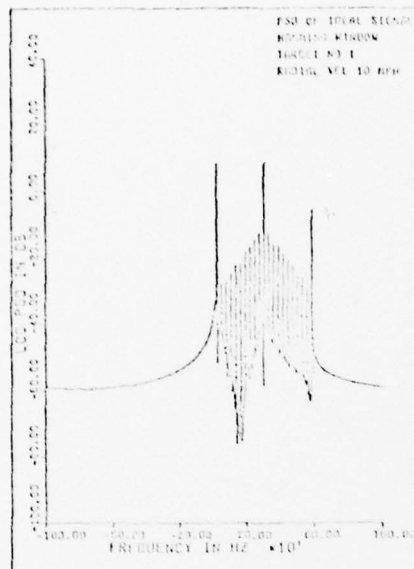
Fig 13. (Continued)

Preprocessing the Data Base. As previously described, the data base used in this experiment consisted of the radar returns from various types of vehicles being driven through the beams of two types of radars at different azimuth angles. Both the in-phase and quadrature components of the received signal were recorded on analog tape for the first data base and only one component for the other.

Initially, the signatures were sampled at twice the bandwidth of the tape and resulting spectra were analyzed. It soon became apparent, consistent with observation 4 about PAM targets, that the doppler spread was essentially confined to the band between 0 and $2\omega_D$. Accordingly, the data was sampled at a rate compatible with the highest target radial velocity.

Because of the physical considerations previously detailed, the pattern space chosen was the space of short-time discrete Fourier amplitude spectra. An FFT integration period (corresponding to a 1024 point FFT) was selected that was short enough to approximate quasi-stationarity of the time series and long enough to ensure sufficient spectral resolution. As discussed in the previous section, several time windows were considered for spectral smoothing, and it was concluded that the Hanning window provided adequate performance at acceptable computational cost.

A FORTRAN computer program called WRENCH was written to perform the required preprocessing on the Wright-Patterson AFB CDC CYBER-74 computer. A flow chart of the program is shown in Fig 14. First the program reads a time record from the input, digital, time domain tape. If the input tape has been completely processed, the program stops. Otherwise, the time data is multiplied by a Hanning window and FFTed. The FFT is, of course, an efficient implementation of the DFT:

56
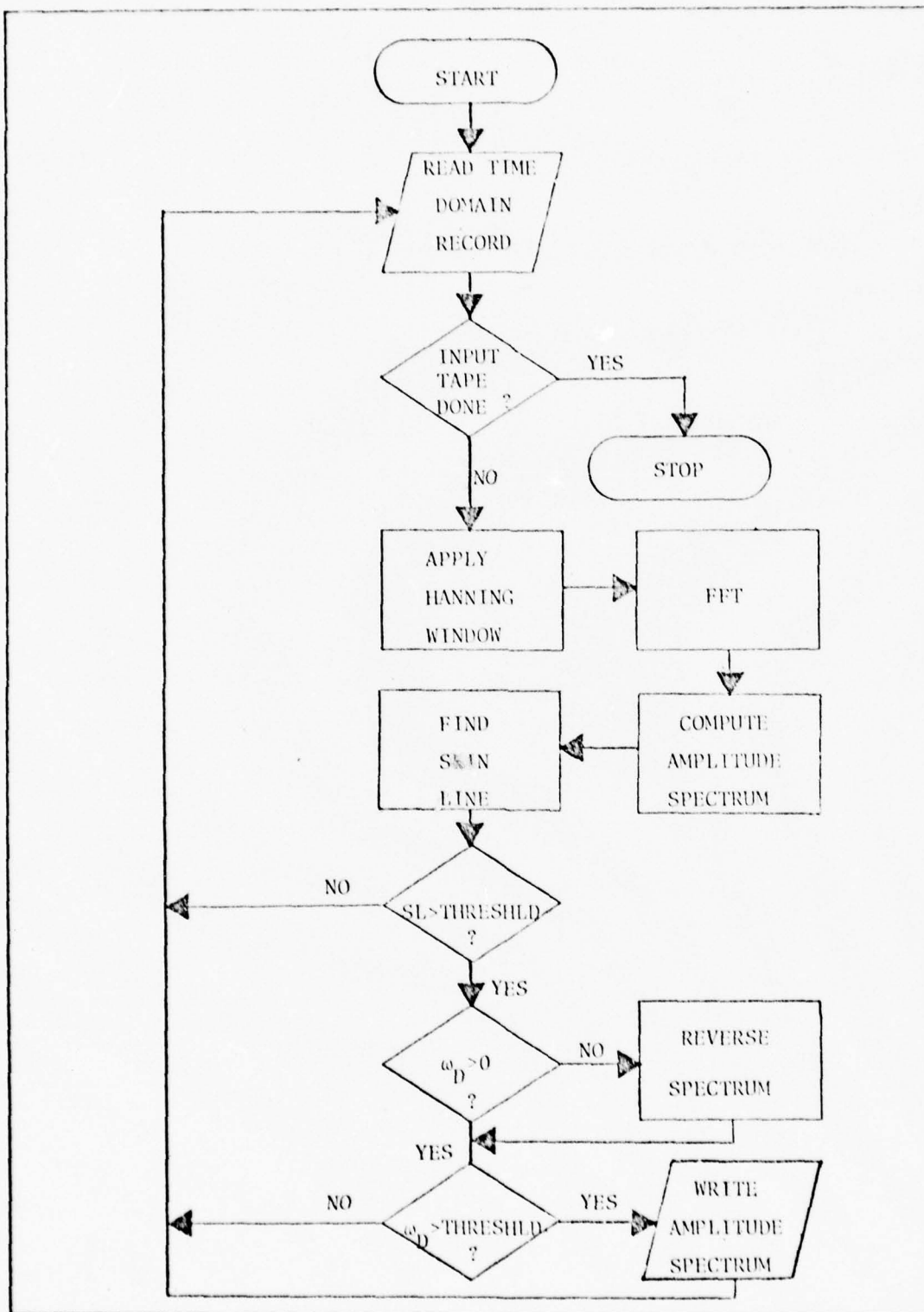
Fig 14. Flow Chart for Program WRENCH

57

$$S(k) = \sum_{n=0}^{N-1} s(n) \exp(-j2\pi kn/N), k=1,2,\ldots N \qquad (55)$$

The particular FFT algorithm used here is FFT2 from the International Mathematical and Statistical Library (Ref 35). FFT2 utilizes a modification of the Singleton (Ref 62) version of the Cooley-Tukey FFT algorithm. It requires the standard $N \log_2 N$ basic sets of operations. After the complex spectrum has been computed, only the amplitude spectrum is retained.

$$|S(k)| = \{[\text{Re } S(k)]^2 + [\text{Im } S(k)]^2\}^{1/2}, k=1,2,\ldots N \qquad (56)$$

At this point the skin line of the signature is found. A very simple approach was taken; the peak signal other than ground clutter was called the skin line. This procedure found the actual skin line in over ninety nine percent of the sample spectra; however, for a few samples of Target 4, at zero degree azimuth, one of the sideband spikes was of greater magnitude than the skin line. Such a sample spectrum is depicted in Fig 15, when the skin line is at 350 Hz and the peak signal is at 500 Hz. Since this type of phenomenon was so rare, and because the feature extraction procedure adopted did not use the fine structure of the individual spectra, it was concluded that no refinement of the skin line extraction procedure was required.

Subsequent to the skin line identification, three thresholding operations were applied to the skin line magnitude and frequency. First the skin line magnitude was compared against an empirically determined threshold. If the magnitude was less than the threshold, the target was considered not to be in the range or azimuth gate of the radar, and that sample was rejected. If the sample spectrum passed the first threshold test, the skin line doppler was checked for positiveness. If it was negative, the spectrum was reversed

$$S_R(k) = S(N+1-k) , k = 1,2,\ldots N \qquad (57)$$

58
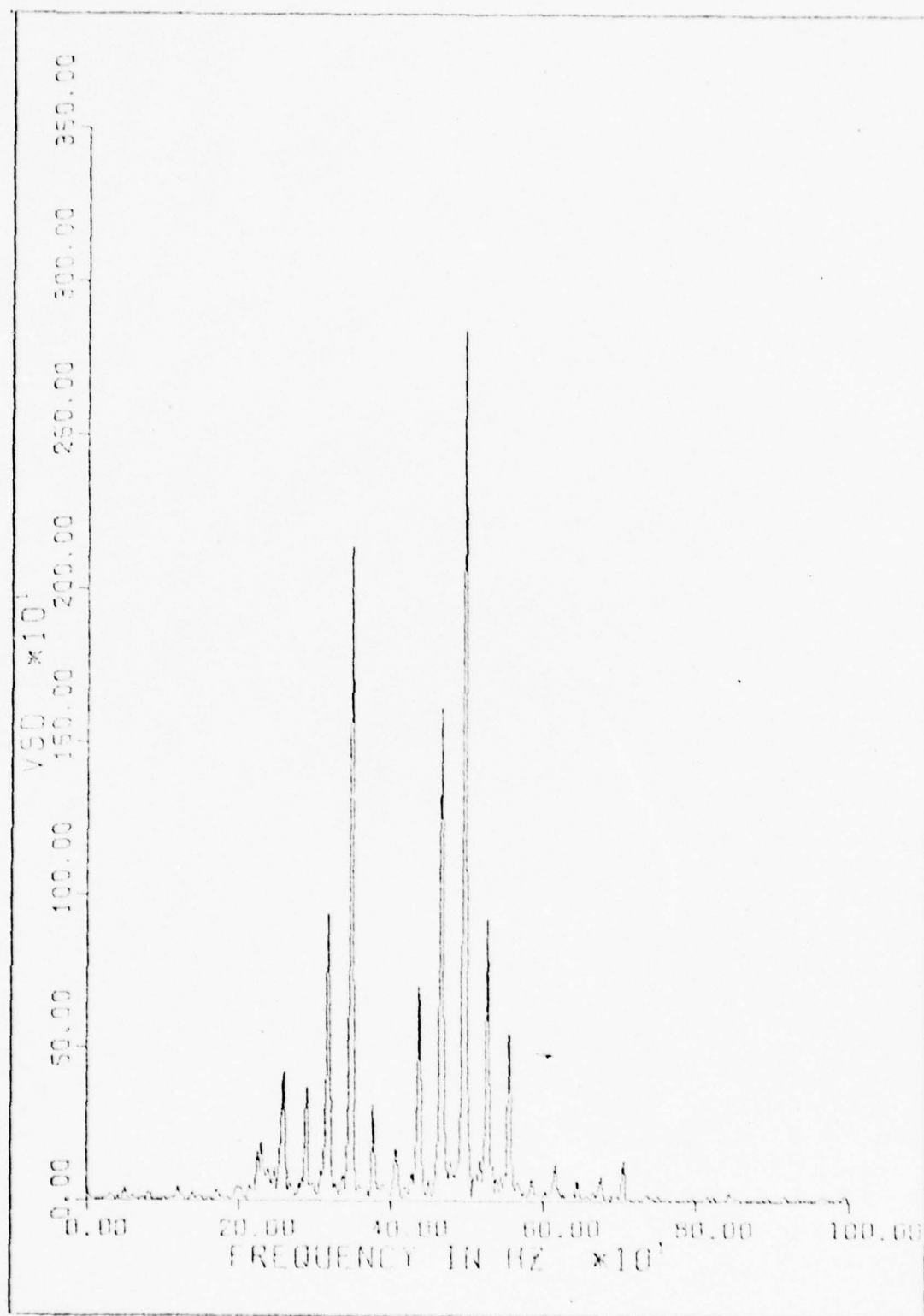
Fig 15.   Amplitude Spectrum of Target 4 at Zero Degree Azimuth

to insure that the skin line doppler was positive for all spectra. After this operation, the negative half of the amplitude spectrum was discarded. Then the skin line frequency was compared to a frequency threshold that corresponded to a radial velocity of about two miles per hour. If the skin line frequency was below that threshold, there was not enough doppler spread to extract useful features, and the sample was rejected. If the sample spectrum passed this final test, the positive half of the amplitude spectrum was written on the output tape.

In summary, the actual preprocessing consisted of a series of transformations and tests applied to the data. First, the data was prefiltered and digitized. Then it was windowed and transformed via FFT, and only the amplitude spectrum was saved. The skin line was extracted and threshold in frequency and magnitude. The resulting output was the positive half, amplitude spectrum of the selected samples.

The resulting baseband amplitude spectrogram for Target 1 at zero degree azimuth is shown in Fig 16. The series of high amplitude spikes at about 300 Hz is the skin line of the vehicle. The lower amplitude, periodic spikes are due to running gear modulation. There is no significant signal beyond the second harmonic of the skin line. It is apparent that the amplitude of any particular spike varies with time in a random fashion. There is some time correlation, however, and a certain degree of frequency correlation exists among neighboring spikes. Both the time and frequency correlation are due to the scattering geometry and how fast it is changing. The geometry, in turn, depends on the terrain and the steering of the vehicle.

Fig 17 depicts the spectrogram of the same vehicle at 135 degrees azimuth aspect. As seems to be true in general, the time coherrence of the sideband spikes for this target is not as great as azimuths other than
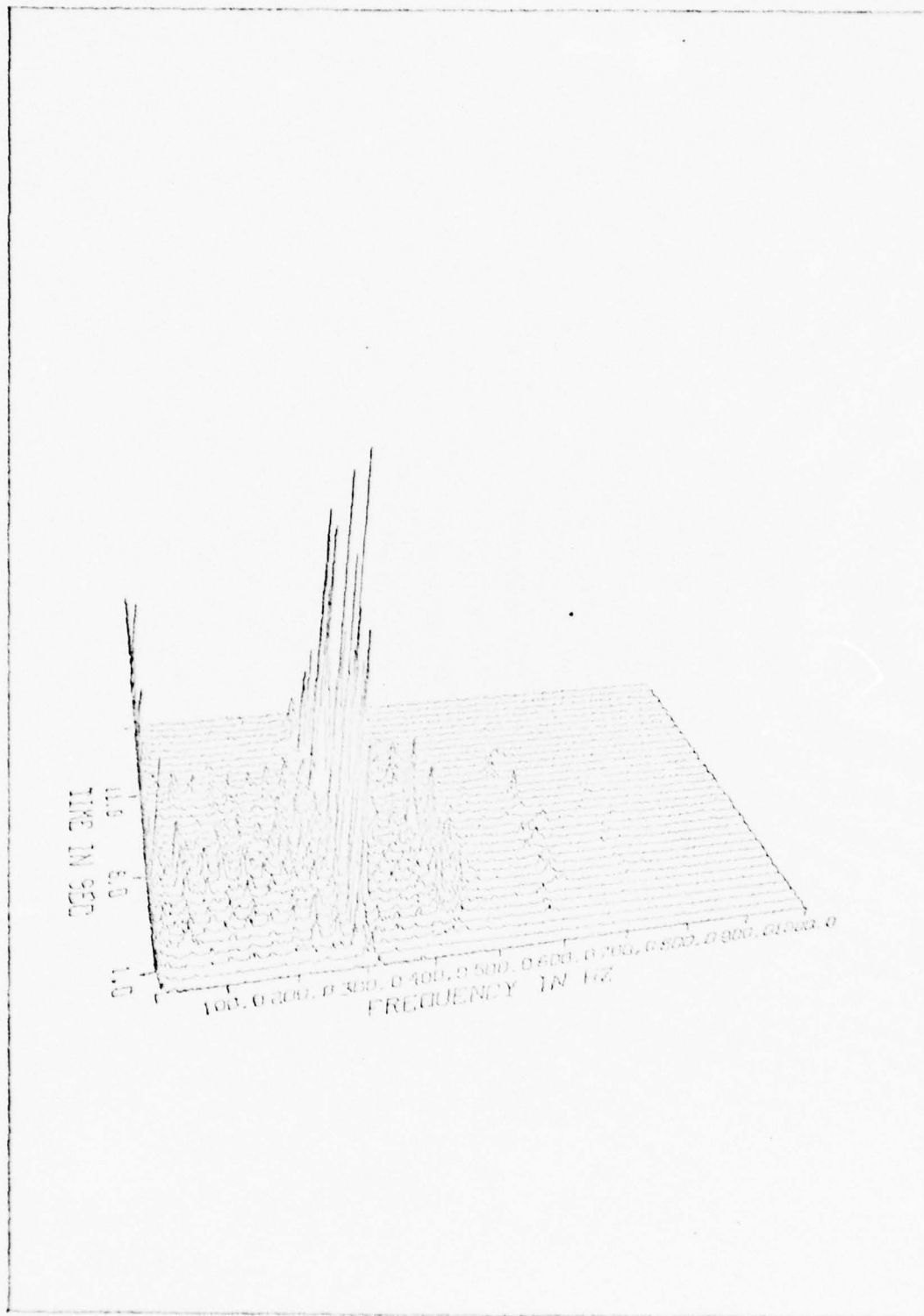
Fig 16.   Spectrogram of Target 1 at Zero Degree Azimuth

zero or 180 degrees. This is probably due to the fact that at those two angles the running gear rotation produces specular flashes. During the period of this spectrogram, the vehicle is accelerating into the range gate and then decelerating out of it.

The spectrogram signature of Target 2 is shown in Fig 18, with the vehicle traveling through the range gate at 315 degrees aspect. The tendency is for this type of target to have a lower level of side band modulation than that of Target 1. These three dimensional spectrograms with hidden lines were made with the Display Integrated Software System and Plotting Language (DISSPLA) (Ref 34).

## Feature Extraction

Feature Extraction may be formally defined as a transformation from the pattern space to the feature space

$$A : G \rightarrow H \qquad (58)$$

where A is the feature extraction transformation, G is the pattern space, and H is the feature space. Typically G and H are both Euclidean spaces with G being of higher dimensionality than H. In any case, H is always a finite dimensional space with each coordinate representing a different feature.

The feature extraction transformation should be considered to be a filter that removes irrelevant information and retains information that is pertinent to the classification problem. In an experimental context, where the engineer is given a data base upon which to design a classifier, feature extraction implies that artifacts in the data must be ignored. Several such artifacts were observed in the data used in this research. As an example of one such artifact consider a radar, with an automatic gain control (AGC), having leakage from the power supply. If the pattern space is the short-time Fourier spectrum of the radar signal, the leakage
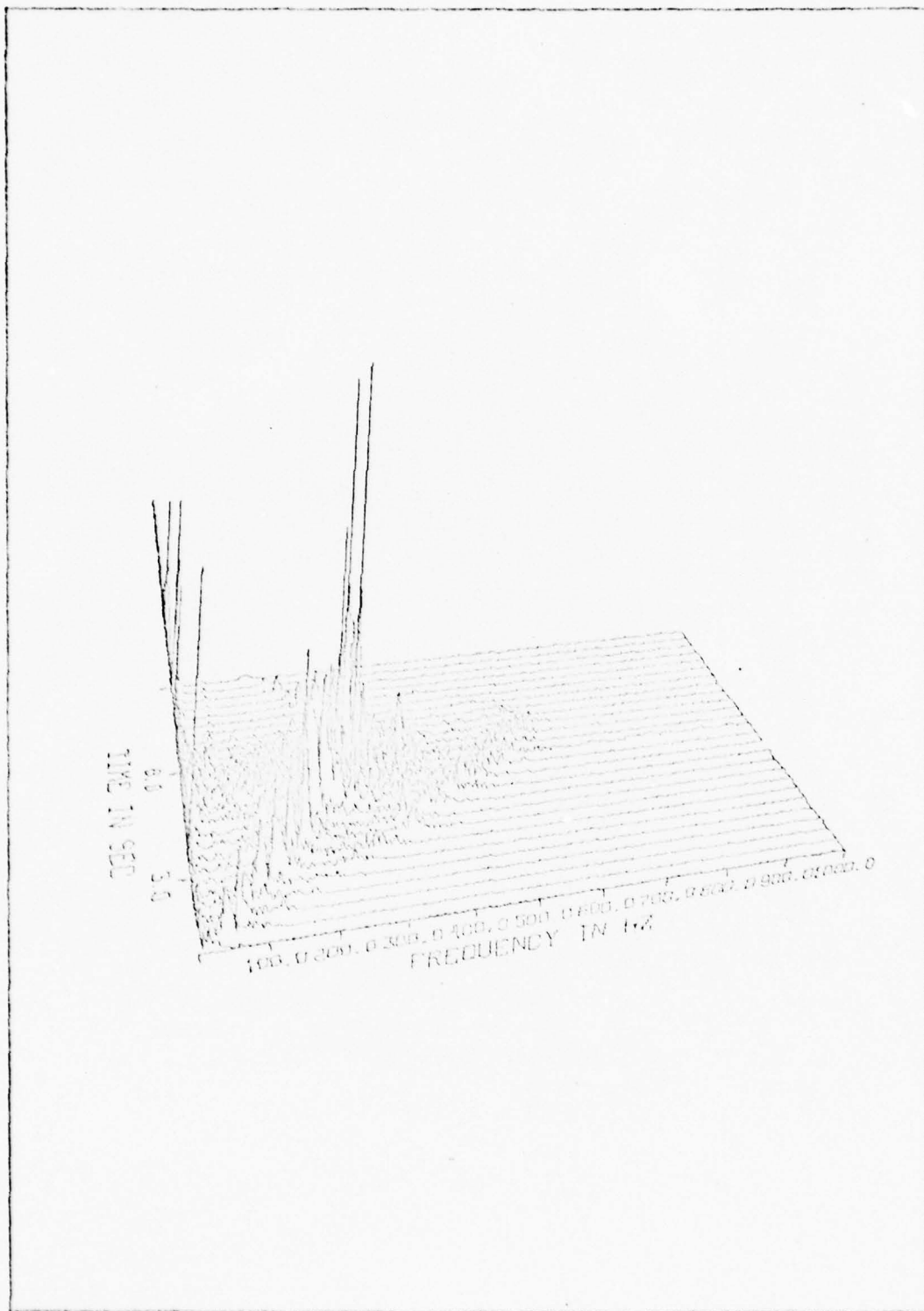
Fig 17. Spectrogram of Target 1 at 135 Degrees Azimuth

63
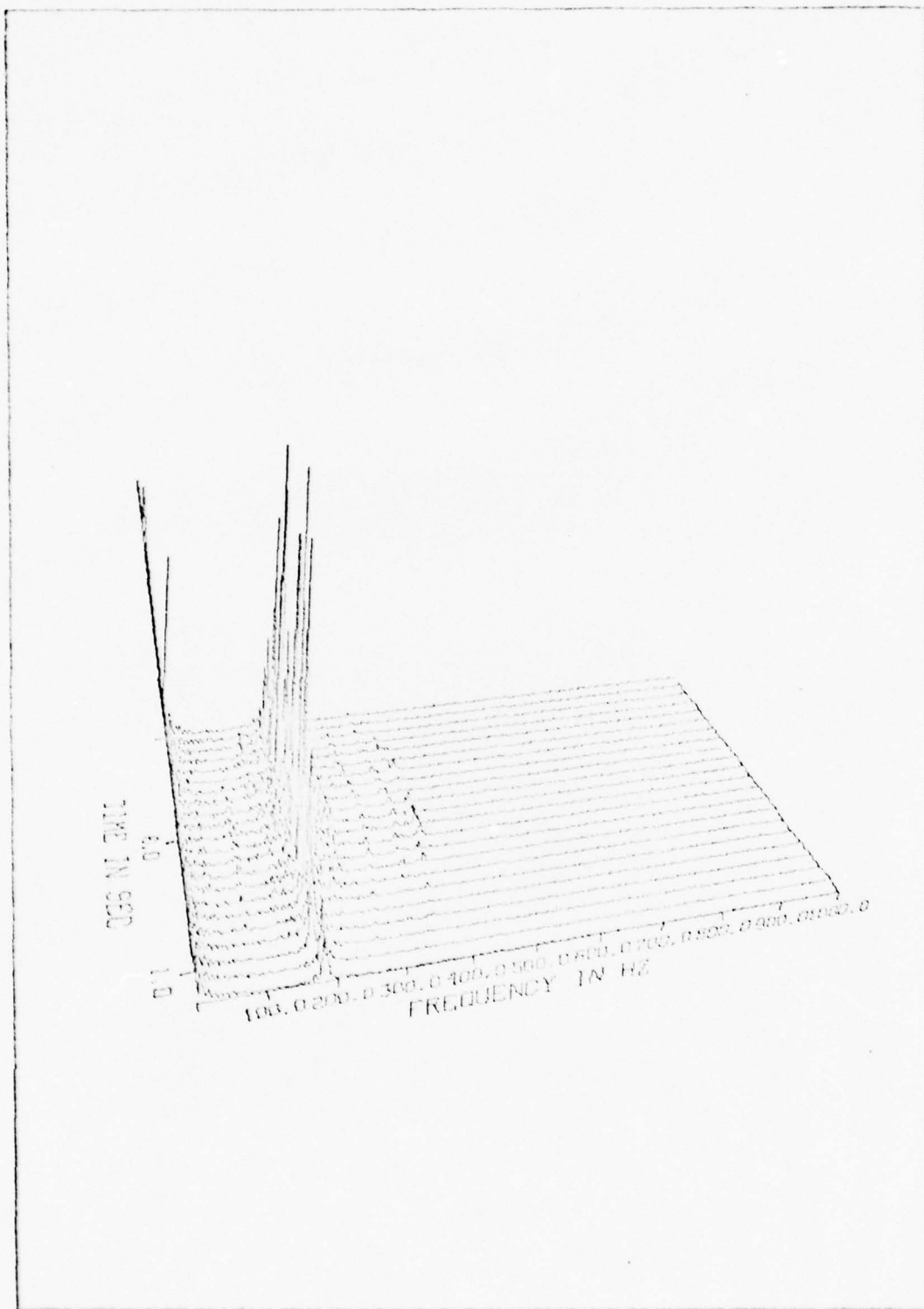
Fig 18. Spectrogram of Target 2 at 315 Degrees Azimuth

will be manifest as a spike at the frequency of the power supply. The AGC will cause the amplitude of the power supply spike to be inversely related to the target strength. Thus the amplitude of the leakage Fourier coefficient might be a useful indication of input signal strength, but it would not be reliable as a feature for distinguishing between different types of targets, since signal strength varies with many more parameters than just target type.

Similarly, if the radar does not have an AGC, strong signals may cause nonlinear distortion in the mixers, resulting in targets with strong harmonics of the skin line. If, as is the case with the targets considered in this research, certain targets tend to possess strong second harmonics, and the data is contaminated by intermodulation harmonics, this second harmonic may be a useful feature, but perhaps, it must be weighted rather lightly.

Because of the high dimensionality of the pattern space ($E^{512}$ initially), and the parametric nonstationarity of the signature, classification in the space of amplitude spectra is not feasible. Therefore, it is desirable to extract features which are somewhat invariant with target parameters.

Some features that could be most useful for discrimination between specific types of vehicles are suggested by the fine structure of the short-time spectra. For example, if the spacing between the sideband spikes could be determined, the period of the amplitude modulating function could be extracted. Then, if the vehicle true velocity could be accurately estimated, the ratio of amplitude modulation period to vehicle velocity could be a reliable feature. Unfortunately, as indicated in the spectrograms, the sideband spikes are not distinct at other than zero or 180 degrees azimuth.

An attempt was made to highlight the formants by integrating over several spectral observations. The procedure consisted of considering several time-consecutive spectral signatures at the same azimuth angle. Each observation was radial velocity normalized, since the skin line frequency may change considerably from one integration period to the next. Then, the normalized signatures were simply summed, with the hope that the coherent signal would be emphasized, while any noise, being incoherent, would not.

This procedure worked failrly well at zero and 180 degrees azimuth, as shown by Fig 19, where the formants were already rather apparent. In Fig 19, it may be noted that the lower side band spikes, between zero and 500 Hz, are quite distinct. Unfortunately, the technique did not emphasize the formants at other azimuths, as in apparent from Fig 20. This is probably due to the lack of specular flashes from the rotating running gear at headings other than zero or 180 degrees. Because of these poor results, this scheme was abandoned; however, it could possibly be made workable by a more refined normalization and interpolation scheme than was used here. A better physical model that more closely simulates the scattering characteristics at other than cardinal headings would provide greater insight and might indicate what modifications might make an integration scheme practicable.

The cepstral signature of the targets seemed an obvious choice for extracting the period of the amplitude modulation, since a series of evenly spaced spectral spikes will integrate into a single large spike in the cepstrum at the period of the spacing. The cepstrum, first described by Bogert and Tukey (Ref 5), is formally defined as the power spectrum of the log power spectrum.

$$C(t) = \left| F\{\log |S(\omega)|^2\} \right|^2 \tag{59}$$

66
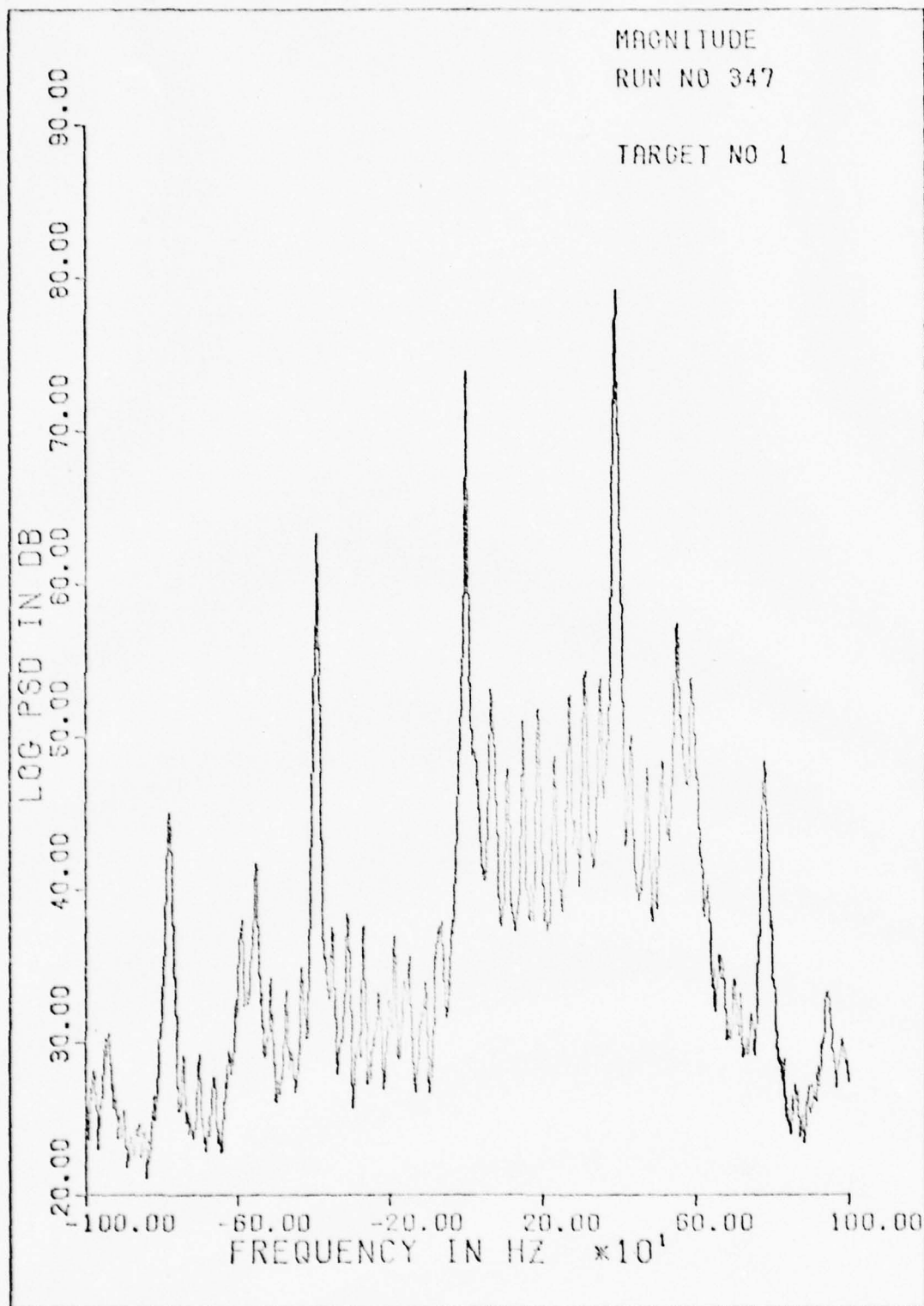
Fig 19. Mean Spectral Signature of Target 1 at Zero Degree Azimuth
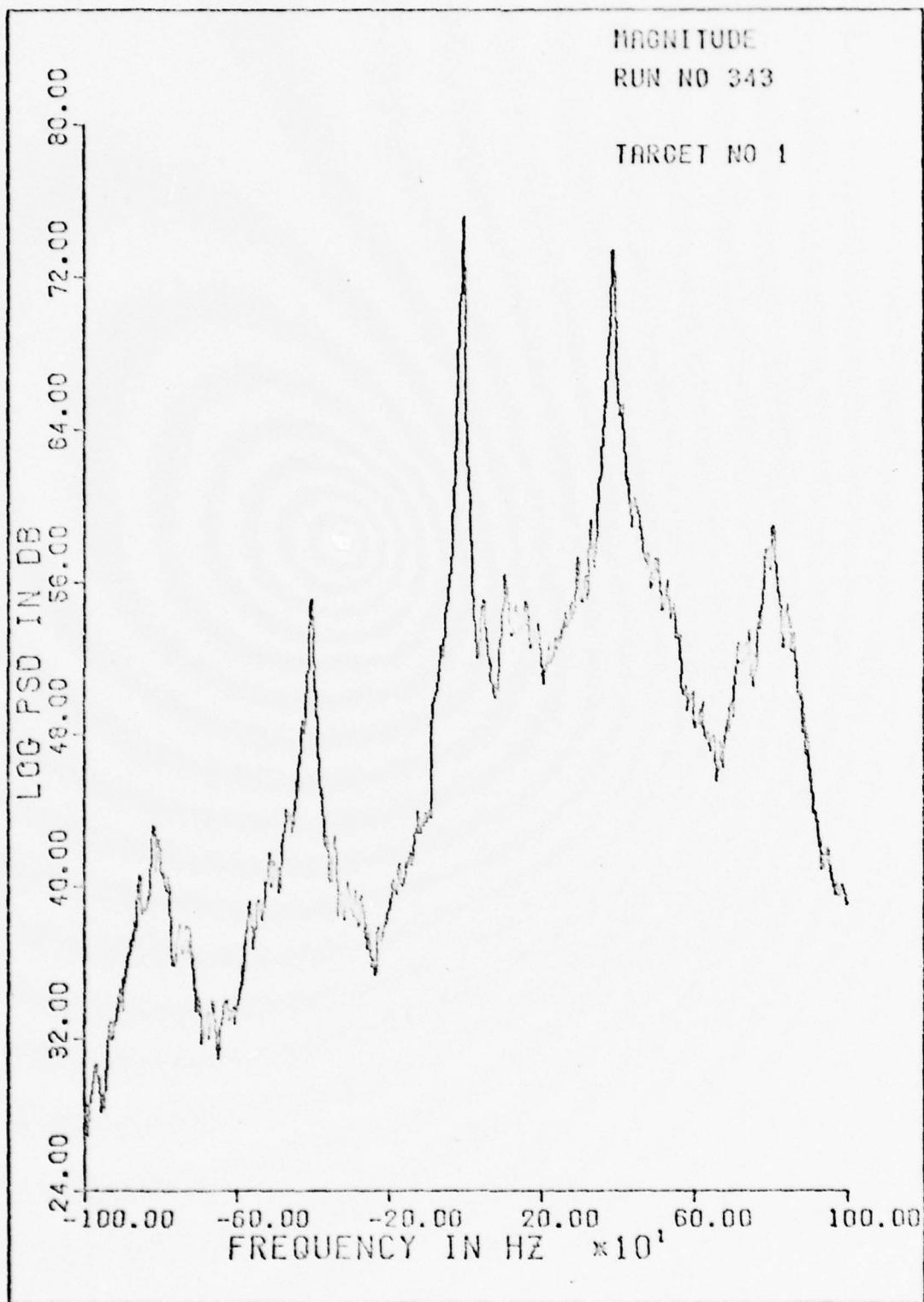
Fig 20. Mean Spectral Signature of Target 1 at 225 Degrees Azimuth

68

A recent tutorial paper on the subject was published by Childers (Ref 10), or for an introduction to homomorphic signal processing, a generalization of cepstral techniques, the reader may wish to consult Oppenheim and Schafer (Ref 51).

The cepstra for a number of observations were computed and the corresponding plots were examined. The results of this experiment were not unlike those of the previous one. The amplitude modulation period could easily be extracted for Target 1 at zero and 180 degrees azimuth angles but could not generally be determined at other azimuths. In Fig 21, the modulation period is represented by the distinct line at about 0.03 seconds, while in Fig 22 no distinct modulation spike is in evidence. The failure of the cepstrum to extract the modulation period at the other azimuths is again due to the lack of distinct formants. Because of these results, the cepstrum was abandoned as a potential source of reliable, aspect invariant features.

Since the attempts to use the fine structure of the individual signatures met with so little success, it was decided to extract features that represented gross characteristics of the signatures. In some ways, this rejection of the fine structure in favor of more gross characteristics is reminiscent of other researchers' attempts to find the Gestalt of processes to be recognized. For example, Radoy (Ref 55) used the low spatial frequency components of the images of alphanumeric characters, with some success, as features in an optical character recognition scheme. Also, researchers at Ohio State University (Ref 41) in their aircraft identification studies have enjoyed greater success utilizing low radar frequencies rather than higher ones. In both cases, the low frequency (gross) information is being retained, and the high frequency (fine detail) information is being discarded. In the optical
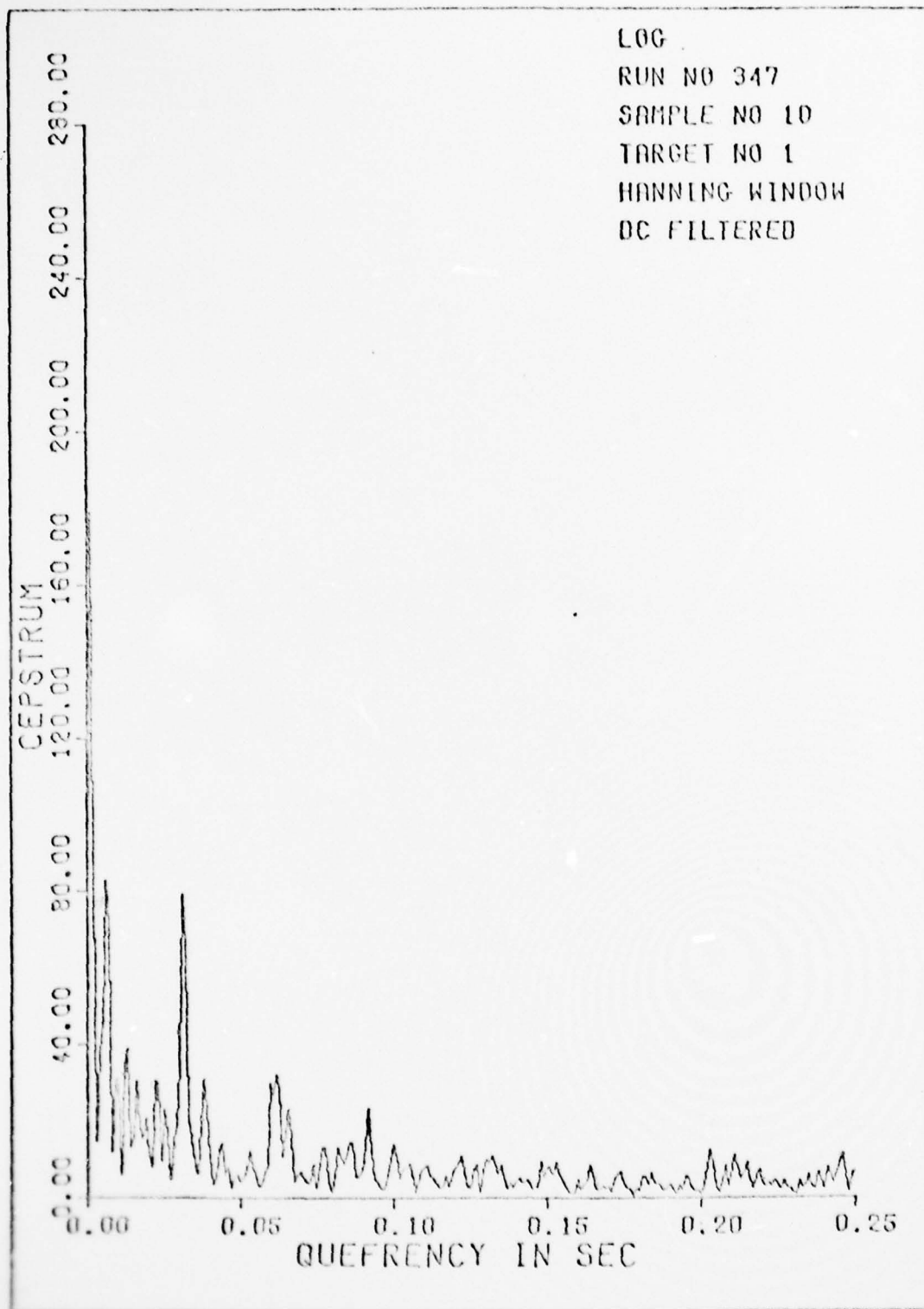
Fig 21. Cepstrum of Target 1 at Zero Degree Azimuth
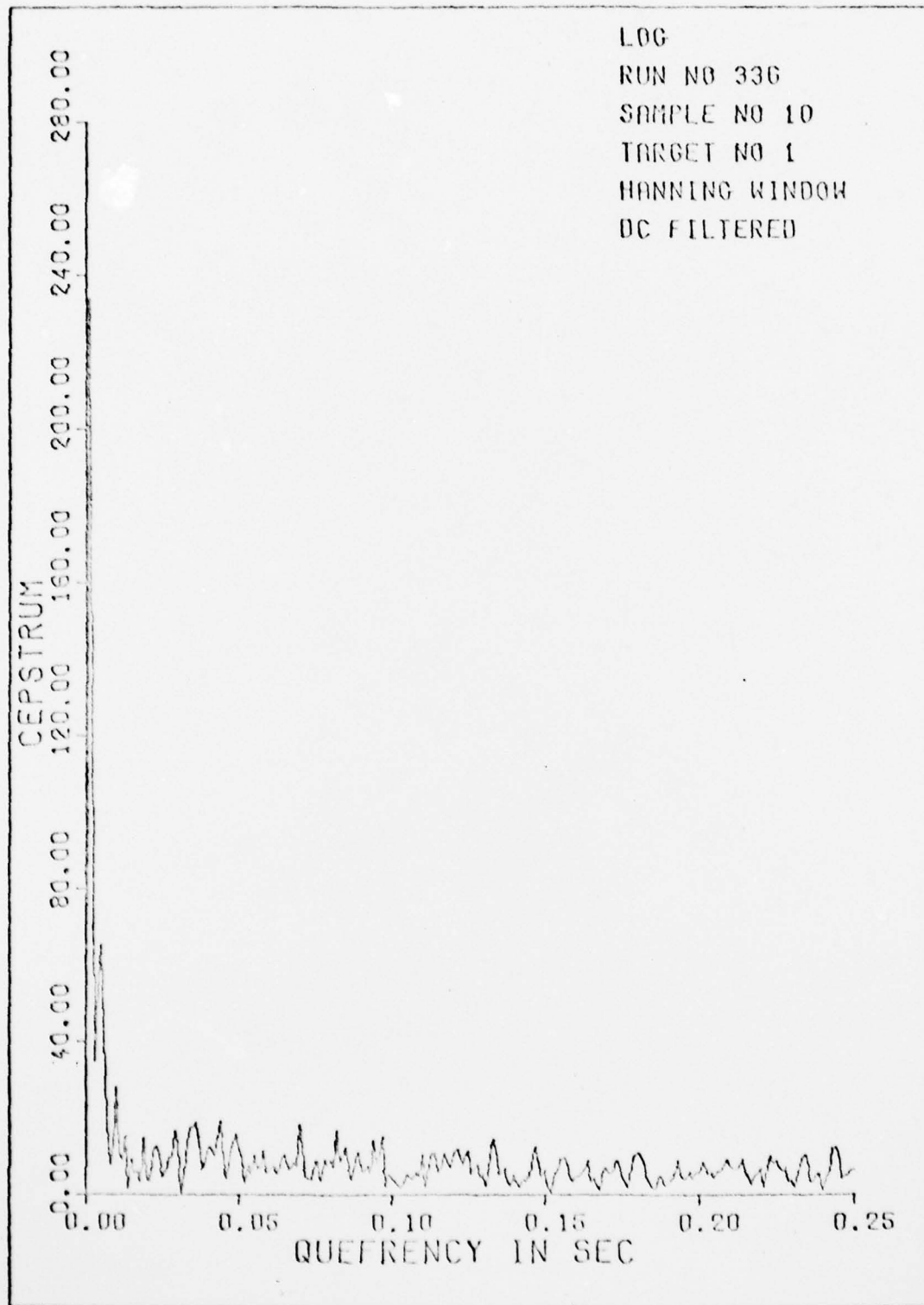
Fig 22. Cepstrum of Target 1 at 135 Degrees Azimuth

character recognition research, it was found that retention of the high frequency components tended to confuse the issue. Too much information was available, and the high frequency information only revealed the fine detail of the structure, while the low frequency components contained the gross form information. The aircraft identification problem is analgous. The low frequenc, components are relatively aspect invariant and contain information about the gross shape of the aircraft while the high frequency information reveals the finer detail such as the tail structure, engine nacelles, etc.

The first step in the adaptive feature extraction algorithm used here is to parse the input spectrum into six bands as depicted in Fig 23. Since the radar antenna is stationary during the data acquisition period, the ground clutter is concentrated near the carrier frequency. Band $B_1$ includes all of the very low frequency components. The information in this region is discarded, since the clutter dominates the signal here. Next, the skin line is found, and a narrow frequency region, $B_3$, is defined centered at the skin line frequency. Band $B_5$, of the same width as, $B_3$, is centered around the second harmonic of the skin line. The lower side band, $B_2$, consists of all frequency components between $B_1$ and $B_3$. The upper side band, $B_4$, is defined as all frequencies between $B_3$ and $B_5$. Then the noise band, $B_6$, consisting of all frequencies higher than those in $B_5$, is ignored.

Two types of features were extracted from the data, the first kind being termed a spectral ratio feature and the second a shape factor. The form of the spectral ratio features was suggested by the norms on various Banach spaces (Ref 42). Extracting the peak signal from a frequency band corresponds to the norm on the space of continuous signals $C[a,b]$:

$$\| S(\omega) \|_{C[a,b]} = \max_{a \le \omega \le b} |S(\omega)| \tag{60}$$
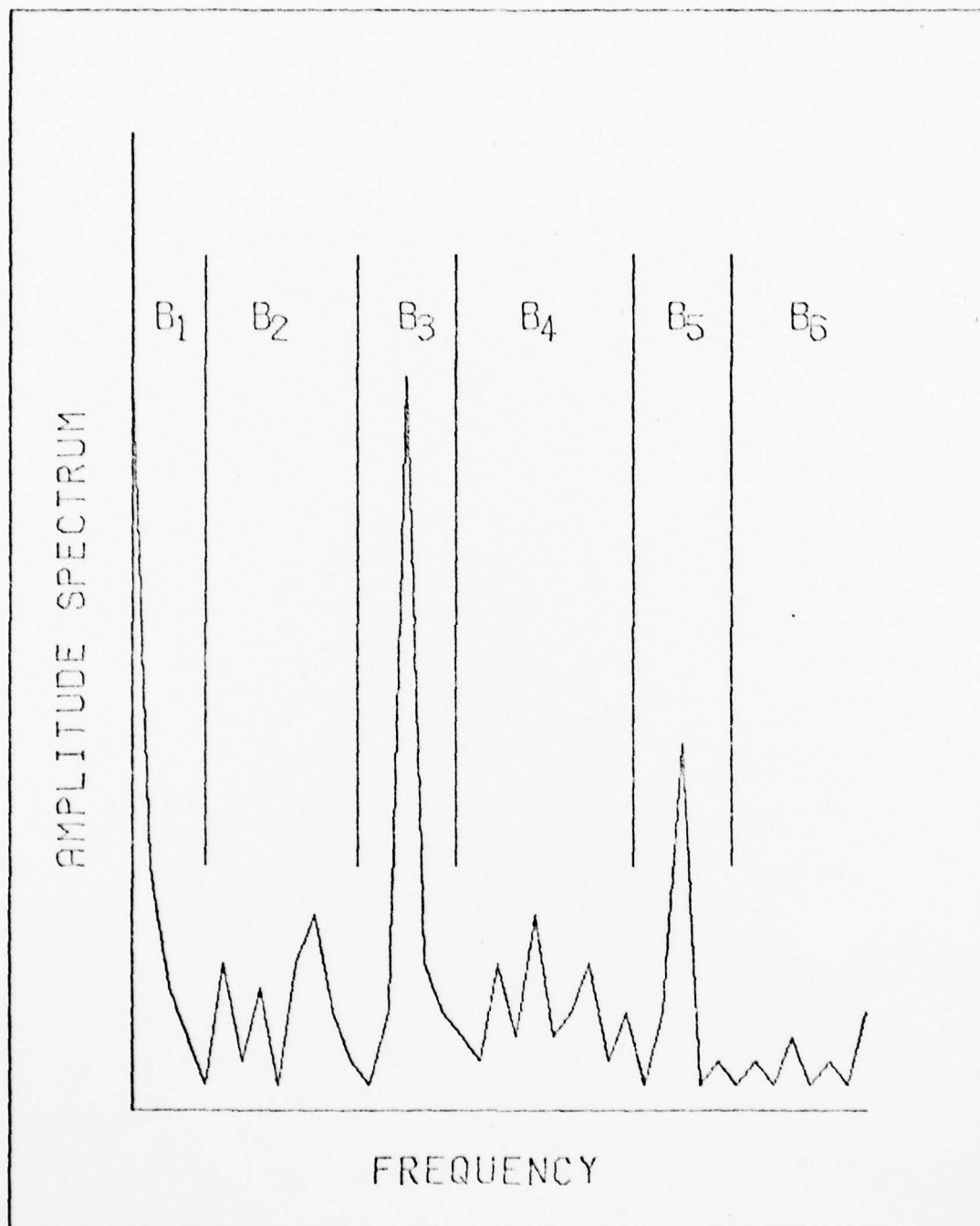
Fig 23.   Spectral Band Structure; where $B_1$=Clutter Band, $B_2$=Lower Side Band, $B_3$=Skin Line Band, $B_4$=Upper Side Band, $B_5$=Second Harmonic Band, and $B_6$=Noise Band

Another quantity of interest, the signal voltage integrated over a given band is derived from the norm on the space of absolutely integrable functions $L_1[a,b]$:

$$\| S(\omega) \|_{L_1[a,b]} = \int_a^b | S(\omega) | d\omega \qquad (61)$$

The signal energy in a specified band coincides with the norm on the space of square-integrable functions $L_2[a,b]$:

$$\| S(\omega) \|_{L_2[a,b]} = [\int_a^b | S(\omega) |^2 \, d\omega]^{1/2} \qquad (62)$$

The final quantity of interest is the total variation (T.V.) within a given band, which is a part of the norm on the space of functions of bounded variation $BV[a,b]$:

$$T.V.[S(\omega)] = \int_a^b | dS(\omega) | \qquad (63)$$

Although the operations listed above are to be performed on continuous signals, the spectral signatures have been discretized, and thus discrete versions of the operations were performed. Also, the actual features are further normalized and averaged as described below.

The first 34 features could be termed spectral ratio features, since they consist of the ratios of characteristics from different bands. The use of these specific features was motivated by both the PAM model and an examination of the spectra of the actual data. Several mathematical quantities that must be calculated to evaluate these features are presented here. One of the first quantities of interest is the peak signal in a given band:

$$P(i) = \max_{n \in B_i}[S(n)] \qquad (64)$$

where $S(n)$ represents the discrete short-time amplitude (voltage) spectrum. The peak signal in the side bands and the second harmonic band appeared to contain class discriminating information. Another attribute of importance is total signal in each band:

$$T(i) = \sum_{n \epsilon B_i} S(n) \tag{65}$$

The total signal and the mean signal in the various sidebands were used as features. The width of each band is calculated according to

$$W(i) = \underset{n \epsilon B_i}{\text{Max}} (n) - \underset{n \epsilon B_i}{\text{Min}} (n) + 1 \tag{66}$$

The band energy is described by

$$E(i) = \sum_{n \epsilon B_i} [S(n)]^2 . \tag{67}$$

Finally, the total variation in each band is calculated by

$$V(i) = \sum_{n \epsilon B_i} |S(n+1) - S(n)| \tag{68}$$

Total variation in the upper and lower sideband was used as a feature to distinguish between the case when the energy is smeared throughout the sidebands and the case when the sideband energy is concentrated in distinct formants. If one of the quantities is to be computed over multiple bands, it is represented, for example, as

$$T(i,j) = \sum_{n \epsilon B_i \cup B_j} S(n) \tag{69}$$

The actual spectral ratio features used are shown in Table I. It may be noted that in each case the features have been normalized by skin line voltage or energy to reduce signal strength variation due to target range or aspect changes or different receiver attenuation settings.

75

Six additional features are computed that could be termed shape factors, since they are functions of moments of the spectrum. These shape factors contain information about the gross, global structure of each spectrum. The motivation for using these features stems from the well known fact that any probability density function is completely specified if all of its moments are known. Furthermore, the greatest amount of information is typically found in the lower order moments. Four mathematical functions are defined that are used in the computation of these shape factors. The first function is the frequency component number of the peak signal in a band:

$$R(i) = n \text{ such that } S(n)=P(i) \tag{70}$$

The mean for a band is defined as

$$M(i) = [\sum_{n \varepsilon B_i} n\, S(n)]/T(i) \tag{71}$$

The variance for a band is

$$A(i) = [\sum_{n \varepsilon B_i} n^2\, S(n)]/T(i)-[M(i)]^2 \tag{72}$$

and the skewness is

$$C(i) = [\sum_{n \varepsilon B_i} n^3\, S(n)]/T(i)-3M(i)\, A(i)+5[M(i)]^3 \tag{73}$$

The shape factor features are the last six defined in Table 1. The shape factors have all been normalized by the skin line frequency to remove shape differences due to changes in radial velocity.

Feature Selection

After the designer has extracted a set of potentially useful features from the preprocessed signals of interest, he is frequently faced with the specification of a further dimensionality reducing transformation.

Table I. Features

| Symbol | Name | Definition |
|---|---|---|
| F(1) | Peak 2 | $P(2)/P(3)$ |
| F(2) | Total 2 | $T(2)/P(3)$ |
| F(3) | Mean 2 | $F(2)/W(2)$ |
| F(4) | Peak energy 2 | $[F(1)]^2$ |
| F(5) | Total energy 2 | $E(2)/[P(3)]^2$ |
| F(6) | Mean energy 2 | $F(5)/W(2)$ |
| F(7) | Total variation 2 | $V(2)/P(3)$ |
| F(8) | Mean variation 2 | $F(7)/W(2)$ |
| F(9) | Peak 4 | $P(4)/P(3)$ |
| F(10) | Total 4 | $T(4)/P(3)$ |
| F(11) | Mean 4 | $F(10)/W(4)$ |
| F(12) | Peak energy 4 | $[F(9)]^2$ |
| F(13) | Total energy 4 | $E(4)/[P(3)]^2$ |
| F(14) | Mean energy 4 | $F(13)/W(4)$ |
| F(15) | Total variation 4 | $V(4)/P(3)$ |
| F(16) | Mean variation 4 | $F(15)/W(2)$ |
| F(17) | Peak 2,4 | $Max[F(1),F(7)]$ |
| F(18) | Total 2,4 | $E(2)+E(10)$ |
| F(19) | Mean 2,4 | $F(18)/[W(2)+W(4)]$ |
| F(20) | Peak energy 2,4 | $Max[F(4),F(12)]$ |
| F(21) | Total energy 2,4 | $F(5)+F(13)$ |
| F(22) | Mean energy 2,4 | $F(21)/[W(2)+W(4)]$ |
| F(23) | Total variation 2,4 | $F(7)+F(15)$ |
| F(24) | Mean variation 2,4 | $F(23)/[W(2)+W(4)]$ |
| F(25) | Peak 5 | $P(5)/P(3)$ |
| F(26) | Peak energy 5 | $[F(25)]^2$ |
| F(27) | Peak 2,4,5 | $Max[F(17),F(25)]$ |
| F(28) | Total 2,4,5 | $F(18)+T(5)/P(3)$ |
| F(29) | Mean 2,4,5 | $F(28)/[W(2)+W(4)+W(5)]$ |
| F(30) | Peak energy 2,4,5 | $Max[F(20),F(26)]$ |
| F(31) | Total energy 2,4,5 | $F(21)+E(5)/[P(3)]^2$ |
| F(32) | Mean energy 2,4,5 | $F(32)/[W(2)+W(4)+W(5)]$ |
| F(33) | Total variation 2,4,5 | $F(23)+V(5)/P(3)$ |
| F(34) | Mean variation 2,4,5 | $F(33)/[W(2)+W(4)+W(5)]$ |
| F(35) | Mean difference 2,3,4 | $|M(2,3,4)-R(3)|/R(3)$ |
| F(36) | Standard deviation 2,3,4 | $[A(2,3,4)]^{1/2}/R(3)$ |
| F(37) | Skewness 2,3,4 | $[C(2,3,4)]^{1/3}/R(3)$ |
| F(38) | Mean difference 2,4 | $|M(2,4)-R(3)|/R(3)$ |
| F(39) | Standard deviation 2,4 | $[A(2,4)]^{1/2}/R(3)$ |
| F(40) | Skewness 2,4 | $[C(,24)]^{1/3}/R(3)$ |

i.e. feature selection. Typically, more features are extracted than can be conveniently used in the classification process. The primary limitation on the number of features that may be used results from the finite sample size of the training set.

If an excessive number of features is used for a given set of observations, overtraining on the design set will occur. The relationship between the number of sample observations available and the optimal number of features to be used is not well understood, even though there is an extensive literature available on the subject. The overtraining will be manifest in one of two ways: either an excessively optimistic evaluation of the selected features' usefulness or an actual decline in performance over that achieved with a smaller number of features.

Foley (Ref 23) has reported the results of an experiment in which he arbitrarily separated samples arising from an artifically generated, single, multivariate uniform density into two classes. By selecting a large number n of features in relation to the number k of samples per class, he could obtain good separation of the data using the Fisher linear discriminant, even though all of the data was generated from the same probability density function. For a ratio of $k/n = 0.36$, he obtained perfect separation. Foley concluded that, on the basis of his experiments, the ratio $k/n$ should be greater than or equal to three to avoid this type of anomaly.

The second type of manifestation of overtraining, and the one that is frequently reported in experimental pattern recognition exercises, is an increase in classifier performance as a function of n out to a certain number followed by a decrease as more features are added. The number of features at which peak performance occurs is referred to as the optimal computational complexity by Chandrasekaran (Ref 8). This phenomenon does

78

not occur for Bayes classification with known densities.

In order to avoid overtraining and to keep on-line feature extraction and classification costs at a minimum, the designer is faced with the problem of how to select the best m features out of the n that were extracted from the design set data. It would be easy enough to estimate the performance of each feature individually (given the form of the classifier to be used) and rank order the features based on their performance. However, if the n extracted features were so ordered, the m individually highest performing features would not necessarily comprise the set of m optimal features (Ref 14). Cover and Van Campenhout (Ref 16) have shown, further, that there exist jointly Gaussian probability density functions on which all possible probability of error orderings can occur among subsets of several measurements subject to a monotonicity constraint. They conclude that "no known nonexhaustive sequential m-measurement procedure is optimal." Exhaustive search is generally out of the question because of the astronomical number of combinations that must be evaluated. If n features have been extracted and it is desired that m of these be selected for classification purposes, then the number of combinations that must be evaluated is equal to

$$\binom{n}{m} = \frac{n!}{m!\,(n-m)!} \tag{74}$$

As Stearns (Ref 64) has noted, typically n is of the order 100 and m is of order 10, then $\binom{100}{10} = 17 \times 10^{12}$.

Fukunaga and Narendra (Ref 26) have advanced a branch and bound algorithm for feature subset selection that restricts the domain of the exhaustive search. The technique consists of the search of a tree whose nodes represent the sets of features included in the evaluation. The root

79

of the tree represents the set of all features, and each of its n

successors represent the exhaustive set with a different feature removed.

Each succeeding set of nodes represent sets reduced by one feature

over their predecessor. The nodes with no successors represent the sets

of m features. The savings in the top-down search procedure arise

because no successors must be evaluated, if performance falls below

an already established lower bound. As Fukunaga notes, the scheme

only works if the performance criterion is monotonic on the number of

features used. Performance improvement of a subset of features over

the superset from which they are drawn cannot be allowed. Thus, this fea-

ture selection technique will work well when performance measures

that presuppose exact knowledge of the statistics of the processes, such

as Battacharyya distance or divergence, are used. However, its optima-

lity will not be retained for practical problems where probability of

error is the performance criterion, and an optimal computational

complexity exists.

Since the only optimal feature selection technique for a practical

pattern recognition problem is exhaustive search which is computationally

exorbitant, heuristic search procedures are required. Mucciardi and

Gose (Ref 46) have provided a comparison of seven heuristic feature

selection techniques. Stearns (Ref 64) has proposed a "plus m ,

take away n" search algorithm that is intuitively appealing. Others, e.g.

Chang (Ref 9), have applied dynamic programming to the feature selection

problem. Stepwise discriminant analysis, as described in the next section,

was chosen as the feature selection technique to be used for the present

problem, because it results in a simple classification rule. Also it

is relatively robust as long as the data is unimodal by classes.

## Stepwise Discriminant Analysis

Stepwise discriminant analysis (SDA) is a technique of feature selection and evaluation. From a training set of extracted features, SDA will select the most useful (in a specified sense) features and evaluate the discriminating power of those features using linear discriminant functions. There is no reason to believe that SDA is an optimum procedure, in any analytical sense, for the problem at hand, nor is it optimum for almost any problem to which it is applied. Although SDA is generally suboptimal for selecting features, the optimum technique is too expensive in terms of computation to consider. The use of SDA has been widely reported in the literature with good results, for example, by Mohn (Ref 45).

Besides performing feature selection, SDA effects classification using linear discriminant functions. Nilsson, in his classic book (Ref 49), details the theory of linear discriminant functions for pattern recognition. It is well known that linear discriminant functions are not optimal for classification except for certain types of class conditional probability density functions, including jointly Gaussian with equal covariance matrices (Ref 18:29). Use of linear discriminant functions may be justified on the basis that if acceptable performance results, then such discriminants are useful, though suboptimal.

The goal of SDA is to find a linear discriminant function that maximizes the ratio of between-class scatter to within-class scatter. Since it is an iterative process, adding or deleting a feature on each pass, it is referred to as stepwise. The motivation for the procedure is based on the classical work of Fisher (Ref 21).

The input to the SDA algorithm is a set of n-dimensional feature vectors $X_1, X_2, \ldots X_k$ assigned to two or more classes: $\omega_1, \omega_2, \ldots, \omega_m$. The

81

technique is a supervised learning method since the class labels are known for each feature vector; however, it is nonparametric, since the class conditional probability density functions are not known. The output of SDA is a set of linear discriminant functions that optimizes a separability criterion.

The data are real and three types of statistics are computed:

1. The overall sample mean vector

$$\bar{X} = \frac{1}{k} \sum_{i=1}^{k} X_i \tag{75}$$

2. The class sample mean vectors

$$\bar{X}_i = \frac{1}{k_i} \sum_{X \in \omega_i} X \quad , i=1,2,\ldots m \tag{76}$$

where $k_i$ equals the number of samples assigned to $\omega_i$.

3. The unbiased estimates of the class variances for each feature

$$\sigma_{ij}^2 = \frac{1}{k_i-1} \sum_{p=1}^{k_i} (x_{pj} - \bar{x}_{ij})^2 \quad , \quad \begin{matrix} j=1,2,\ldots,n \\ i=1,2,\ldots,m \end{matrix} \tag{77}$$

where i refers to class number and j to the feature number.

At each step of the analysis, two types of overall scatter matrices are computed. First, the individual class scatter matrices are evaluated

$$S_i = \sum_{X \in \omega_i} (X-\bar{X}_i)(X-\bar{X}_i)^T \quad , i=1,2,\ldots m \tag{78}$$

from which the (pooled) within-class scatter matrix is formed

$$S_W = \sum_{i=1}^{m} S_i \tag{79}$$

The within-class scatter matrix, as its name implies, gives some measure of the total within-class scatter. The total scatter matrix

$$S_T = \sum_{i=1}^{k} (X_i-\bar{X})(X_i-\bar{X})^T \tag{80}$$

indicates the total scatter of all of the samples. For the scalar case, it may be noted that the within-class scatter is the sum of the normalized class variances, and the total scatter is the overall normalized variance. It may also be observed that

$$S_T = \sum_{i=1}^{m} \sum_{X \epsilon \omega_i} (X - \bar{X}_i + \bar{X}_i - \bar{X})(X - \bar{X}_i + \bar{X}_i - \bar{X})^T$$

$$= \sum_{i=1}^{m} \sum_{X \epsilon \omega_i} (X - \bar{X}_i)(X - \bar{X}_i)^T + \sum_{i=1}^{m} \sum_{X \epsilon \omega_i} (\bar{X}_i - \bar{X})(\bar{X}_i - \bar{X})^T$$

$$= S_W + \sum_{i=1}^{m} k_i (\bar{X}_i - \bar{X})(\bar{X}_i - \bar{X})^T \tag{81}$$

The second term in Eq (81) is frequently termed the between-class scatter matrix $S_B$ (Ref 18:119), and thus total scatter in the sum of the within-class scatter and the between-class scatter

$$S_T = S_W + S_B \tag{82}$$

The method of selecting those features to be added or deleted at each step of the analysis is based on the ratio of within-class scatter to total scatter. This quantity, which is the ratio of the determinants of the within-class and total scatter matrices, is referred to as Wilks' $\Lambda$-criterion (Ref 36:77)

$$\Lambda(Y) = \frac{\det S_W (Y)}{\det S_T (Y)} \tag{83}$$

where Y is the vector of features that have been selected and det represents determinant. Wilks' $\Lambda$-criterion has values between zero and unity, with larger values indicating poor separation, while smaller ones indicate good separation. At each step, the features are divided into two disjoint groups, those that have been selected at some previous step and those that have not. A partial $\Lambda$-statistic is defined

83

$$\Lambda(u,Y) = \frac{\Lambda((Y,u))}{\Lambda(Y)} \tag{84}$$

where u is a feature that has not been selected as yet. This quantity gives a measure of improvement of the expanded feature set over the original set. A corresponding F-statistic is computed

$$F = \frac{k-m-p}{m-1} \cdot \frac{1-\Lambda(u,Y)}{\Lambda(u,Y)} \tag{85}$$

where p is the number of features included in the analysis. The F-statistic is used to control the addition of new features or the deletion of already selected features. The F-statistic is called either the F-to-enter statistic for the entry of u into the set $Y=(y_1,y_2,\ldots y_p)$ or the F-to-remove statistic for the deletion of u from the set $(y_1,y_2,\ldots,y_p,u)$. A new feature will not be added if its F-statistic is below a specified threshold, or an old one will not be deleted if its F-statistic is above a specified threshold. Both thresholds are input parameters in the computer program.

At each step of the analysis, one feature is added or deleted according to the following three rules:

1. Remove the feature with the smallest F-to-remove value unless this value is greater than or equal to the F-to-remove threshold.

2. If it is not possible to remove a feature, add the feature with the highest F-to-enter statistic that is greater than or equal to the F-to enter threshold.

3. If it is not possible to delete or add a feature, the stepping procedure is complete.

After the iterations have ceased, a linear discriminant function is computed for each class:

$$D_i(X) = (X - 1/2\bar{X}_i)\, \Sigma^{-1}\bar{X}_i^T + P(\omega_i), \quad i = 1, 2, \ldots, m \tag{86}$$

where X is an arbitary feature vector composed of the features selected by SDA, and $\bar{X}_i$, is the sample mean vector for those features. The pooled within-class sample covariance matrix in Eq (86) is

$$\Sigma = \frac{S_w}{k - m} \tag{87}$$

for the selected features. The linear discriminant functions defined by Eq (86) are optimal for Gaussian class conditional densities with equal covariance matrices (Ref 18:29).

For a more rigorous and complete presentation of SDA, as well as a flow chart, the reader is referred to Jenrich (Ref 36). The actual computer program used is BMD07M which is contained in the BMD Biomedical Computer Programs Series (Ref 17:214a-214t).

Besides the linear discriminant functions, other outputs from BMD07M include a classification matrix, and posterior probabilities and Mahalanobis distances for each sample and class. Also, a scatter plot of the data is made using the two best variables produced by a Karhunen-Loe've expansion (Ref 25: 226-233) on the pooled covariance matrix.

The primary disadvantage of using SDA for feature selection and estimation of the performance of the features thus selected is that it only uses second order statistics. As long as the experimental data is unimodal for each class, this is not a serious drawback. Frequently, estimates of higher order statistics are rather unreliable for experimental data in any case. Application of SDA to a wide variety of feature selection problems has demonstrated it's robustness. Also, the linear discriminant classifier is simple and cheap, in terms of memory and computation time, to implement.

## Nearest Neighbor Classification Rule

The concept of geometric nearness as a measure of similarity is implicit in all statistical pattern recognition formulations. The nearest neighbor (NN) classification rule, first proposed by Fix and Hodges (Ref 22), elegantly captures this concept. The NN training set consists of a number of d-dimensional feature vectors $\{X_1, X_2, \ldots, X_n\}$ that are specified to belong to two or more classes, $\omega_1, \omega_2, \ldots, \omega_m$. A suitable metric $d(\cdot, \cdot)$ is defined on the space to which the feature vectors belong. A test vector $Y$ of unknown class label is assigned to the class $\omega_c$ such that

$$X_j \in \omega_c \quad \text{where} \quad d(Y, X_j) = \min_{1 \leq i \leq n} d(Y, X_i) \tag{88}$$

The vector of unknown classification is assigned to the class to which its nearest neighbor belongs.

The NN rule has been applied to very diverse classification problems, probably because of its intuitive appeal, its ease of implementation, and its robustness with respect to the underlying statistics of the data. The NN classification rule will yield acceptable performance when other classifiers fail, e.g. when applied to data having multimodal probability density functions. Use of the NN rule requires no knowledge of the underlying statistics of the data, and thus it is termed a nonparametric technique. Because of its structure, the rule can provide clustering information about the data.

Cover and Hart (Ref 15) have shown that the NN rule asymptotic probability of error, $P$ is bounded above by twice the Bayes probability of error, $P^*$. If $P_n(e)$ is the NN classification average probability of error for a design set containing n samples, then the asymptotic error rate is

86

$$P = \lim_{n \to \infty} P_n(e) \qquad (89)$$

Cover and Hart then demonstrate that

$$P \leq P^\star(2 - \frac{m}{m-1} P^\star) \qquad (90)$$

where m is the number of classes. The lower bound on P is obviously the Bayes error rate. This near optimality of the NN rule for large training sets, then, provides further justification for its wide use for classification and feature evaluation.

The expected probability of error for the NN rule is easy to compute and is reliable (assuming the design set is representative) since each training set sample may be considered individually as a test set of one observation. This method of estimating the expected probability of error is a generalization of the U method described by Toussaint (Ref 66), who compares its statistical properties to those of other methods. The procedure is to determine the number of samples in each class that have nearest neighbors outside of the class, divide by the number of samples in the class, multiply by the class prior probability, and sum over all the classes. To express this formally, it is first convenient to define a set membership function for the nearest neighbor of a given feature vector $X_j \in \omega_c$

$$\delta(X_j, X_j') = \begin{cases} 1, & \text{if } X_j' \notin \omega_c \\ 0, & \text{if } X_j' \in \omega_c \end{cases} \qquad (91)$$

where $X_j'$ is the nearest neighbor of $X_j$, i.e.

$$d(X_j, X_j') = \min_{i \neq j} d(X_j, X_i) \qquad (92)$$

Then the expected probability of error is

$$P(e) = \sum_{c=1}^{m} \frac{P(\omega_c)}{n_c} \sum_{X_j \in \omega_c} \delta(X_j, X'_j) \tag{93}$$

where $P(\omega_c)$ is the prior probability of class c and $n_c$ is the number of samples assigned to $\omega_c$. It should be noted that the metric most frequently used with the NN rule, and the one used in this work, is Euclidean distance:

$$d(X,Y) = \| X-Y \|_E = [(X-Y)^T (X-Y)]^{1/2} \tag{94}$$

The greatest drawback to using a NN classifier for on-line pattern recognition is the large amount of computational resources required. Each of the design set vectors must be stored, and all of the distances must be computed for each test vector. To alleviate this computational requirement somewhat, Hart (Ref 33) has introduced the condensed nearest neighbor (CNN) rule. The CNN rule uses an iterative procedure to find only those samples that are near the class boundaries. By retaining only those samples, the CNN classification procedure has the possibility of reducing the on-line computational requirements considerably, without effecting performance significantly.

The k nearest neighbor (KNN) rule is an obvious extension of the NN classification rule. As the name implies, this rule simply assigns a test vector to the class which is most frequently represented among its k nearest neighbors. A simple vote is taken to determine the class label for the test vector. The asymptotic upper bound for the KNN probability of error may not be expressed simply; however, the asymptotic performance is monotonic on k. Duda and Hart (Ref 18:105) provide a chart depicting asymptotic KNN performance as a function of Bayes error for various values of k.

Results

To test the efficacy of the extracted features, several classification

experiments were conducted using SDA and NN classification. In all of
the experiments described in this section, unless otherwise indicated,
the feature vectors consisted of the forty features, previously described,
extracted from a 512 point, positive, amplitude spectrum. Data Base A
was comprised of 620 samples of Target 1, 296 of Target 2, and 471 of
Target 3. Data Base B had an additional 131 samples of Target 1 and 53
of Target 2 plus 83 samples of Target 4, 18 of Target 5 and 65 of Target
6. Table XV in Chapter V summarizes the various experiments conducted.

In the first set of experiments, Data Base A was used with Targets
2 and 3 united into a single class, since they are of similar type. Thus
for this two class problem there were 620 samples in the first class and
767 in the second. Experiment 1 consisted of the application of SDA to
this data base with equal prior class probabilities assigned. Table II
shows which features were selected or removed at each step. The F-
statistic, as defined in the section on SDA, controls the entry or re-
moval of features. The features are as defined in Table I in the section
on feature extraction. As can be seen from this table, F(28), the total
signal in the sidebands including the second harmonic, is the best single
feature. Feature F(21) is the total energy in the lower and upper side-
bands, while F(23) is the total variation in those two bands. Features
F(36) and F(38) are both shape factors. Table III shows the features
included at selected steps, ranked in order of F-statistic, i.e., the
"best" feature is first followed by the remainder in descending order of
usefulness. It is interesting to note that F(28) is a reliable feature
when few features are utilized, but it is of less use when a larger number
is selected. The maximum number of features selected is thirty six rather
than forty, because the SDA algorithm rejected the remaining four as being
too highly correlated with those already included. The average probability

## Table 11. Summary of SDA for Experiment 1

| STEP NUMBER | FEATURE ENTERED | REMOVED | F VALUE TO ENTER OR REMOVE | NUMBER OF VARIABLES INCLUDED |
|---|---|---|---|---|
| 1 | 28 | | 774.33 | 1 |
| 2 | 21 | | 242.98 | 2 |
| 3 | 23 | | 158.10 | 3 |
| 4 | 36 | | 264.41 | 4 |
| 5 | 38 | | 100.76 | 5 |
| 6 | 29 | | 98.18 | 6 |
| 7 | 37 | | 35.84 | 7 |
| 8 | 34 | | 18.52 | 8 |
| 9 | 25 | | 12.44 | 9 |
| 10 | 26 | | 78.00 | 10 |
| 11 | 6 | | 21.32 | 11 |
| 12 | 8 | | 17.71 | 12 |
| 13 | 33 | | 10.02 | 13 |
| 14 | 40 | | 9.43 | 14 |
| 15 | 19 | | 8.79 | 15 |
| 16 | 31 | | 10.80 | 16 |
| 17 | 1 | | 9.66 | 17 |
| 18 | 30 | | 9.87 | 18 |
| 19 | 27 | | 14.07 | 19 |
| 20 | 39 | | 13.77 | 20 |
| 21 | 35 | | 7.26 | 21 |
| 22 | 10 | | 4.24 | 22 |
| 23 | 5 | | 7.88 | 23 |
| 24 | 24 | | 2.51 | 24 |
| 25 | 16 | | 4.07 | 25 |
| 26 | | 38 | 0.01 | 24 |
| 27 | 4 | | 1.26 | 25 |
| 28 | 18 | | 1.18 | 26 |
| 29 | 11 | | 1.01 | 27 |
| 30 | 3 | | 14.68 | 28 |
| 31 | | 23 | 0.01 | 27 |
| 32 | 32 | | 8.67 | 28 |
| 33 | 14 | | 3.51 | 29 |
| 34 | | 28 | 0.01 | 28 |
| 35 | 38 | | 3.30 | 29 |
| 36 | 7 | | 1.84 | 30 |
| 37 | 9 | | 1.37 | 31 |
| 38 | 12 | | 0.19 | 32 |
| 39 | 20 | | 0.20 | 33 |
| 40 | 17 | | 0.63 | 34 |
| 41 | 15 | | 0.06 | 35 |
| 42 | 28 | | 0.01 | 36 |

90

# Table III. Features Included for Selected Steps
## in Experiment 1

| STEP NUMBER | FEATURES INCLUDED, ORDERED ON F |
|---|---|
| 1 | 28 |
| 2 | 28,21 |
| 3 | 28,21,23 |
| 4 | 28,23,36,21 |
| 5 | 28,36,23,38,21 |
| 6 | 28,36,23,38,21,29 |
| 7 | 36,28,23,21,38,29,37 |
| 8 | 36,38,21,28,23,29,36,34 |
| 10 | 36,25,38,26,21,29,28,34,37,23 |
| 15 | 36,21,38,26,25,28,6,37,8,29,33,40,19,23,34 |
| 20 | 36,38,28,29,25,6,30,19,27,37,40,39,26,31,8,33, 1,21,34,23 |
| 25 | 36,28,29,6,30,37,19,25,27,31,40,10,35,39,21, 8,5,26,24,34,16,33,1,23,38 |
| 35 | 36,29,18,25,19,6,33,39,8,37,24,40,16,10,11,5, 32,3,2,7,30,14,26,31,38,34,21,35,1,4 |
| 42 | 36,29,19,39,8,6,24,16,37,25,11,40,3,5,32,14, 38,10,26,35,1,30,34,33,18,27,9,4,20,17,31, 12,7,21,15,28 |

of error as a function of the number of features selected is shown in Fig 24. The average probability of error decreases monotonically from .211 for a single feature to .061 for thirty six features. The monotonic decrease in probability of error demonstrates that the computational complexity of this experiment is greater than thirty six
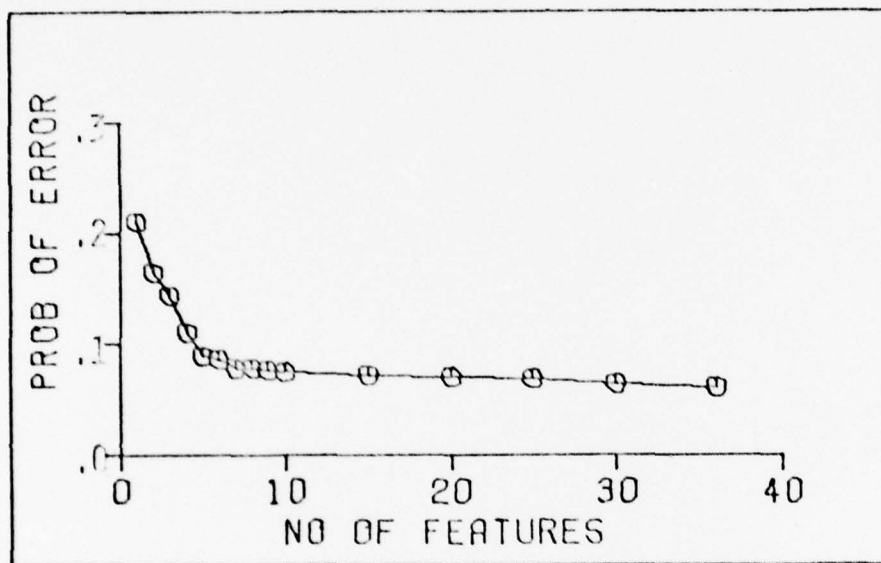


Fig 24. Probability of Error as a Function of the Number of Features Used for Experiment 1

features, which might be expected because of the large size of the design set classes. This figure also shows the diminishing utility of each additional feature. The use of five features reduces the probability of error to less than ten percent. The addition of thirty one more features only lowers the error by a few percentage points.

The thirty six feature SDA classification matrix for Experiment 1 is shown in Table IV. The error bias toward class 1 is caused by several phenomena. Qualitatively speaking, the difference between the sample

Table IV. Confusion Matrix for Experiment 1

|  |  | CLASSIFIED | AS |
|---|---|---|---|
|  |  | 1 | 2 |
| TRUE | 1 | 555 | 65 |
| CLASS | 2 | 14 | 753 |

spectra of the two classes is that those from class 1 tend to have more sideband modulation than those of class 2. When conditions cause class 1 signatures to have less sideband modulation, they are misclassified. As previously noted, the Target 1 modulation spectra at zero and 180 degrees azimuth angles appear different from those at other aspects. This difference is reflected by the fact that twenty seven out of fifty eight samples at those two cardinal aspect angles were misclassified. Another eighteen of the errors occurred with the target backing up. Of the remaining twenty misclassified samples, a large percentage occurred either early or late in the run, indicating the target may have been only marginally within the radar gate. Fine tuning the skin line amplitude threshold could possibly eliminate some of these errors.

The performance cited in this experiment and in subsequent SDA experiments could be criticized as being optimistic due to testing on the training set. However, because of the large size of the training set classes, performance would not be significantly different if a small number of samples were withheld for testing purposes. Furthermore, the nearest neighbor classification experiments did not test on the training set, and the performance was not much different from the SDA results.

A scatter plot for Experiment 1 is depicted in Fig 25. The data is projected onto the best two coordinates arising from the SDA canonical analysis. The asterisks indicate the class means, and the dollar signs

Fig 25.  Scatterplot for Experiment 1

represent two-class overlap. This plot indicates that these two classes
are both essentially unimodal.

In a target identification application, a pattern recognition system
may be designed to withhold its decision until several observations are
made. Consequently, a sequential analysis was performed on the results
from Experiment 1. Because of the nonstationary nature of the signa-
tures, a decision was made on each sample, and then a simple majority
vote was taken to determine the final decision. The results of this
sequential analysis are shown in Fig 26, compared against single and
three nearest neighbor performance. The probability of error decreased
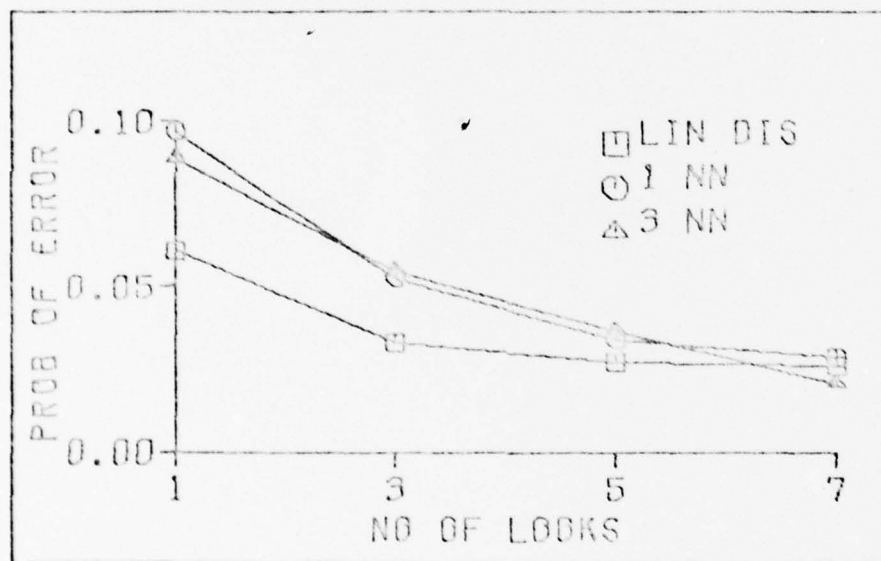


Fig 26. Two Class Performance, Experiments 1, 3, and 4

monotonically from .061 for one observation to .026 for seven for SDA.
When the classification decision was based on seven observations, there were
no class 2 samples misclassified, and all of the errors on class 1 were
at zero or 180 degrees. Since the results for these two cardinal headings
seem to be somewhat exceptional, the performance estimates may actually be
somewhat pessimistic as these headings may be overly represented in the

95

training set. Almost ten percent of the samples were at these two vehicle aspects. The runs with aspects nearest these were 22.5 degrees off, and no problem was encountered with those runs. If the exceptional behavior is assumed to extend for only five degrees either side of the cardinal headings, and all aspects are assumed equally likely, then the cardinal heading runs should only account for about five percent of the total samples. Thus the error rate would only be about half the indicated one.

Experiment 2 was the same as the first except that the Target 1 samples at zero and 180 degrees were included in a separate class. The SDA iterations stopped again after thirty six features had been selected with an overall probability of error of .053 for a single observation. The confusion matrix is shown in Table V, where A represents the Target 1 samples at zero and 180 degrees. The probability

| | | CLASSIFIED | AS | |
|---|---|---|---|---|
| | | A | 1 | 2 |
| TRUE | A | 32 | 2 | 23 |
| CLASS | 1 | 10 | 508 | 44 |
| | 2 | 0 | 7 | 760 |

Table V. Confusion Matrix for Experiment 2

of error given above did not include class 1/class A errors. It is interesting to note that less than one percent of the class 2 targets were misclassified. Also, the vast majority of class A targets that were missidentified were called class 2. The scatterplot for this experiment, shown in Fig 27, provides more insight into the structure of the data.

Experiment 3 consisted of applying the NN classification rule to the data as described in Experiment 1. The average probability of error
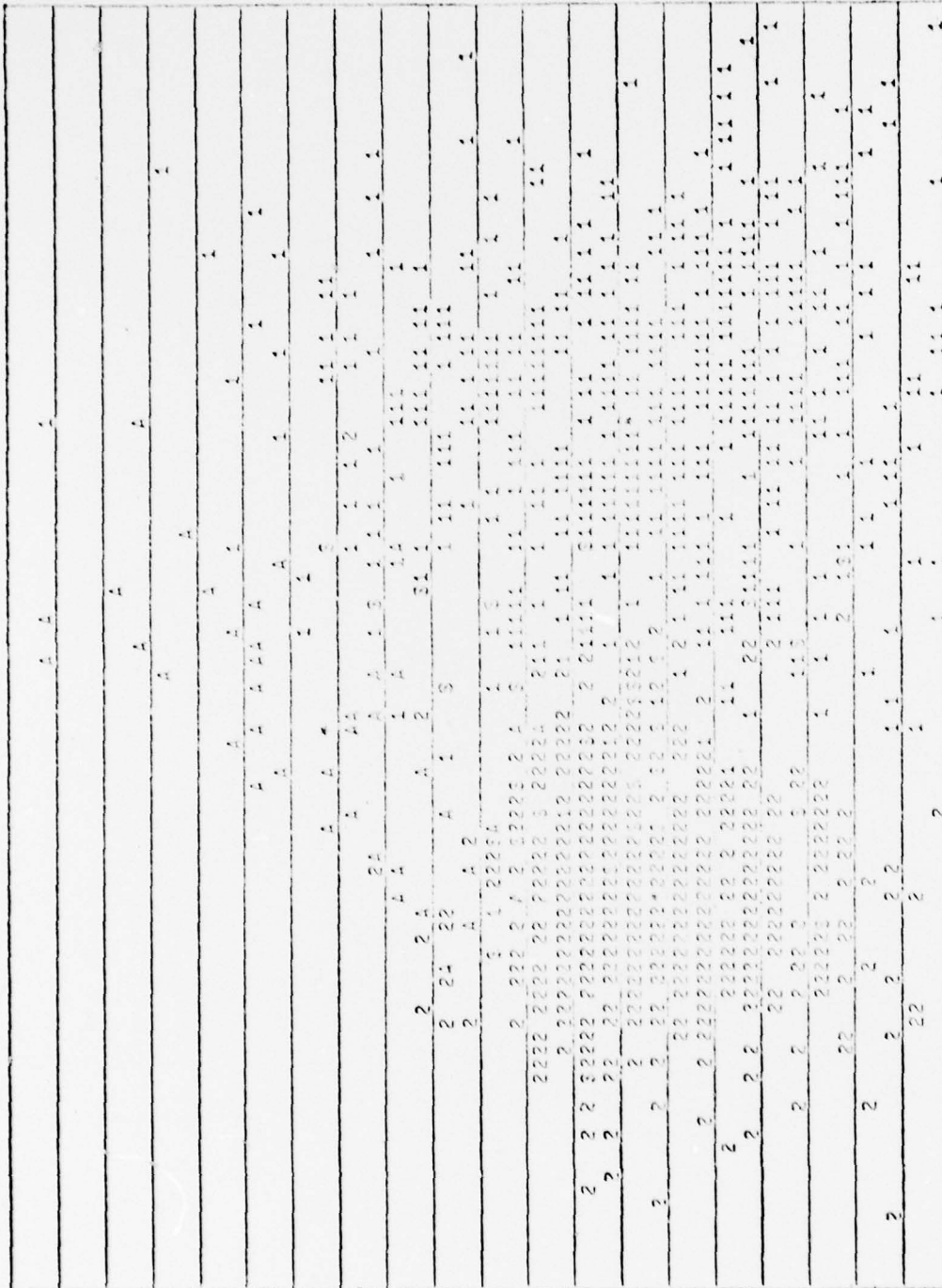
96

Fig 27. Scatter Plot for Experiment 2

vs the number of observations is shown in Fig 26. The probability of
error decreased from .097 for one look to .029 for seven. The NN per-
formance was inferior to that of SDA for one and three looks but was about
the same for five and seven. The confusion matrix is presented in
Table VI with the first number representing one-look performance, and
the second, seven-look. Again, the Target 1 runs at zero and 180 degrees

|        |   | CLASSIFIED | AS |
|--------|---|------------|----|
|        |   | 1          | 2  |
| TRUE   | 1 | 529/588    | 91/32 |
| CLASS  | 2 | 36/5       | 741/762 |

Table VI.   Experiment 3 Confusion Matrix

proved troublesome, causing twenty four percent of the single look errors.
The runs where Target 1 was in reverse caused a significant portion of
the remaining errors, twenty nine percent of the single look and thirty
one percent of the seven look errors.

Experiment 4 was identical to Experiment 3 except a 3NN rule was used
for classification. The performance of the 3NN rule was not significantly
different from that of the single NN rule, as demonstrated in Fig 26. The
3NN rule did, however, require twenty eight percent more computer resources
to run.

Experiment 5 was an application of SDA to the same data base used in
Experiment 1 except that only the upper sideband and second harmonic fea-
tures were used. This included features $F(9)$ through $F(16)$ and $F(25)$ and
$F(26)$. The motivation for this experiment was to determine how the
performance might be degraded by distributed clutter in the lower sideband
that could be experienced by an airborne radar. If the signal to clutter

ratio was poor, the lower sideband spectral ratio features and the shape factors would be useless. The performance for this experiment increased from a probability of error of .128 for a single observation to .067 for seven.

Experiment 6 was an application of SDA to Data Base A with each of the three targets considered as a separate class. Thus class 1 had 620 samples; class 2 had 296; and class 3 had 471. Each class was considered to have equal prior probability. Table VII shows a summary of the SDA iterations. The SDA performance for this three class problem is shown in Fig 28. The relatively poor performance in this experiment is due to



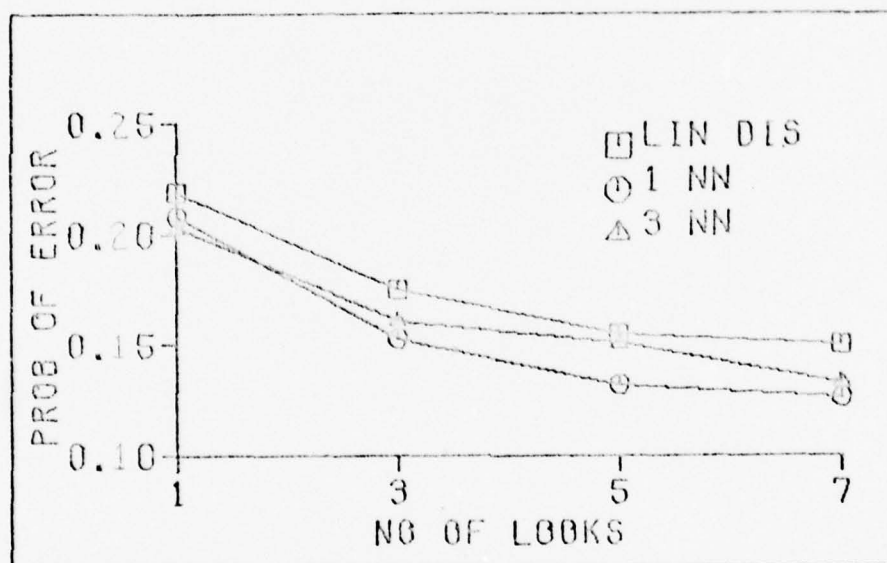Fig 28. Three Class Performance, Experiments 6, 7, and 8

the similarity between the modulation spectra of Targets 2 and 3 as reflected in the confusion matrix, Table VIII, for a single observation. If class 2/class 3 errors are not counted, the performance is about the same as for the previously described two class problem. The scatter plot for Experiment 6 is shown in Fig 29. Table IX reflects the features, ranked

Table VII.  Summary of SDA Iterations for Experiment 6

| STEP NUMBER | FEATURE ENTERED | F VALUE TO ENTER | NUMBER OF VARIABLES INCLUDED | PROBABLITY OF ERROR |
|---|---|---|---|---|
| 1 | 36 | 520.94 | 1 | 0.351 |
| 2 | 37 | 98.08 | 2 | |
| 3 | 4 | 9.63 | 3 | |
| 4 | 8 | 16.91 | 4 | |
| 5 | 22 | 53.29 | 5 | 0.303 |
| 6 | 39 | 26.92 | 6 | |
| 7 | 12 | 4.03 | 7 | |
| 8 | 14 | 5.13 | 8 | |
| 9 | 6 | 20.69 | 9 | |
| 10 | 9 | 24.20 | 10 | 0.265 |
| 11 | 15 | 22.09 | 11 | |
| 12 | 21 | 9.40 | 12 | |
| 13 | 1 | 5.10 | 13 | |
| 14 | 10 | 47.17 | 14 | |
| 15 | 35 | 76.31 | 15 | |
| 16 | 38 | 10.58 | 16 | |
| 17 | 20 | 1.91 | 17 | |
| 18 | 17 | 4.04 | 18 | |
| 19 | 40 | 6.45 | 19 | |
| 20 | 3 | 79.99 | 20 | |
| 21 | 27 | 6.23 | 21 | |
| 22 | 26 | 1.81 | 22 | |
| 23 | 25 | 92.18 | 23 | |
| 24 | 16 | 0.23 | 24 | |
| 25 | 32 | 3.01 | 25 | |
| 26 | 28 | 10.43 | 26 | |
| 27 | 7 | 7.00 | 27 | |
| 28 | 33 | 6.00 | 28 | |
| 29 | 31 | 3.82 | 29 | |
| 30 | 18 | 3.57 | 30 | |
| 31 | 13 | 6.02 | 31 | |
| 32 | 34 | 0.43 | 32 | |
| 33 | 11 | 8.38 | 33 | |
| 34 | 29 | 12.89 | 34 | |
| 35 | 30 | 0.45 | 35 | 0.219 |

Table VIII. Experiment 6 Confusion Matrix

|  |  | CLASSIFIED | | AS |
|---|---|---|---|---|
|  |  | 1 | 2 | 3 |
| TRUE | 1 | 545 | 21 | 54 |
| CLASS | 2 | 2 | 202 | 92 |
|  | 3 | 8 | 95 | 368 |

In descending order of F-statistic, selected by SDA at various steps. Feature F(36), the "standard deviation" of the lower and upper sidebands plus the skinline band, was consistently the top feature for this experiment.

Experiments 7 and S consisted of, respectively, single and three NN classification applied to the same data base used in Experiment 6. The results are summarized in Fig 28 along with those from Experiment 6. In general the NN results were slightly better than SDA on this data base.

Table IX. Features Selected at Various Steps for Experiment 6

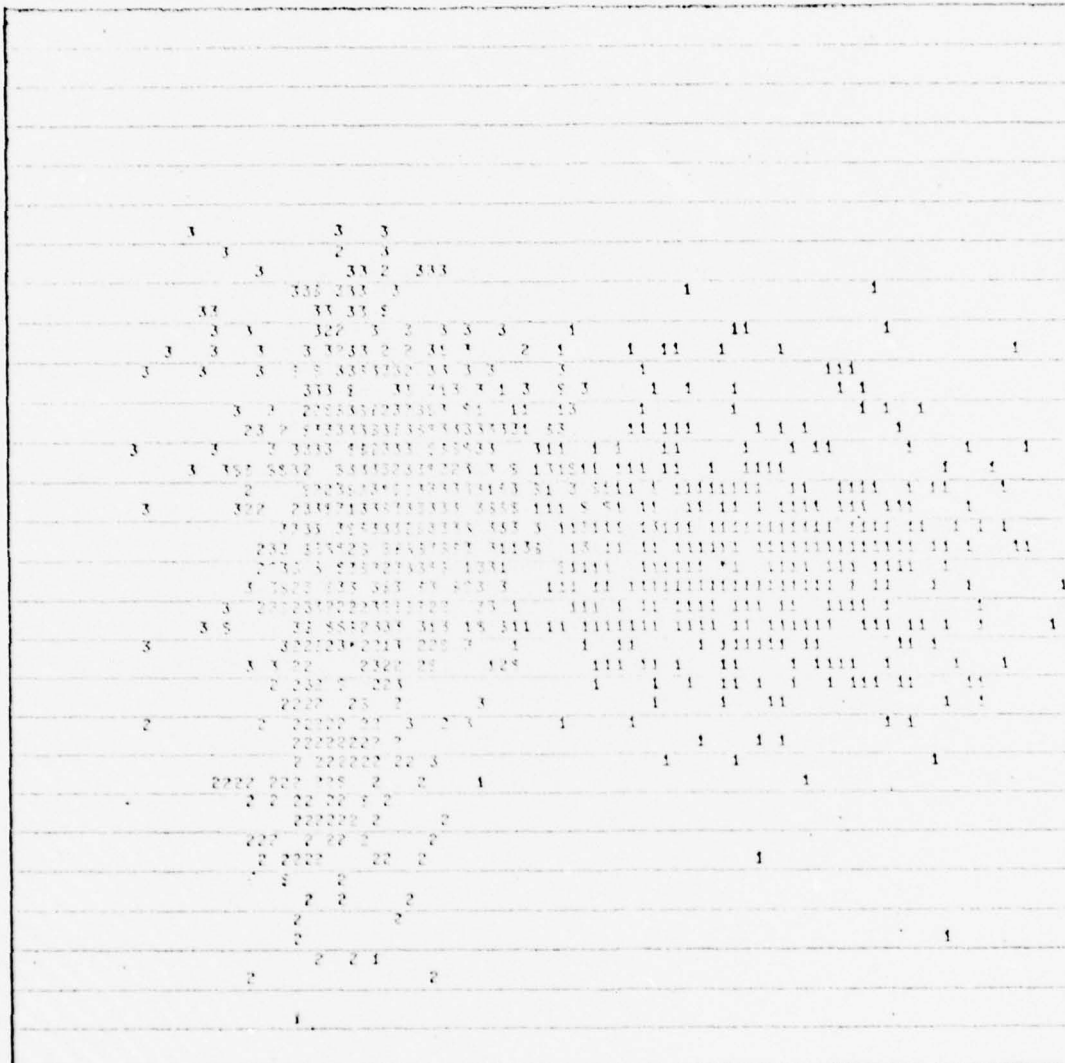| STEP NUMBER | FEATURES INCLUDED, ORDERED ON F |
|---|---|
| 1 | 36 |
| 5 | 36,37,22,8,4 |
| 10 | 36,37,8,4,39,9,6,12,14,22 |
| 20 | 36,3,8,21,10,14,37,22,39,6,35,9 |
|  | 38,4,15,1,12,30,40,17 |
| 30 | 36,39,22,37,35,9,28,38,4,31,14,12, |
|  | 6,40,8,1,25,20,33,17,21,32,18, |
|  | 3,15,10,7,27,26,16 |
| 35 | 36,6,39,22,14,29,10,3,35,13,9,37,38 |
|  | 4,40,1,12,25,32,11,3318,18,17,31, |
|  | 15,20,26,21,28,30,27,7,34,16 |

Fig 29. Scatterplot for Experiment 6

In the remaining series of experiments the training set consisted of Data Base A and Data Base B. This expanded data base consisted of 751 samples of Target 1, 349 of Target 2, 471 of Target 3, 83 of Target 4, 18 of Target 5, and 65 of Target 6. Targets 1 and 4 are of similar type with respect to the structures that cause the modulation, while the remaining four are somewhat similiar.

Experiment 9 consisted of SDA of a two class problem with class 1 comprised of Targets 1 and 4 and class 2 containing the samples from the remaining four targets. A summary of the features selected and the probability of error at certain steps is given in Table X. The table reveals that feature F(36) was once again the best single feature. Table XI shows the features included at selected steps ordered on the F-statistic. A scatter plot for experiment 9 is depicted in Fig 30. The final confusion matrix for this experiment using 34 variables is shown in Table XII for a single look and for seven looks. The probability of error is shown in Fig 31 compared against succeeding experiments.

In Experiment 10 the six targets were each considered as separate classes with prior probabilities equal to the ratio of class samples to total samples. The features selected by SDA are shown in Table XIII. The confusion matrix for a single observation and thirty five features is shown in Table XIV. It is apparent from the confusion matrix that the classifier performance on Target 6 is very poor. As with all Data Base B targets, signatures from this target were only available at fourty five degree increments. At zero and 180 degrees, the Target 6 spectrum had extremely low levels of sideband modulation while at the aspects fourty five degrees off the cardinal headings, the sideband modulation levels were moderately high. Thus the features for Target 6 arise from a bimodal distribution with the cardinal heading signatures being over

Table X.  A Summary of SDA Iterations for Experiment 9

| STEP NUMBER | FEATURE ENTERED | REMOVED | F VALUE TO ENTER OR REMOVE | NUMBER OF FEATURES INCLUDED | PROBABILITY OF ERROR |
|---|---|---|---|---|---|
| 1 | 36 | | 594.16 | 1 | 0.259 |
| 2 | 25 | | 129.03 | 2 | |
| 3 | 37 | | 113.26 | 3 | |
| 4 | 39 | | 73.18 | 4 | |
| 5 | 8 | | 74.74 | 5 | 0.162 |
| 6 | 26 | | 78.35 | 6 | |
| 7 | 3 | | 28.20 | 7 | |
| 8 | 5 | | 56.90 | 8 | |
| 9 | 2 | | 35.57 | 9 | |
| 10 | 7 | | 33.03 | 10 | 0.151 |
| 11 | 4 | | 10.48 | 11 | |
| 12 | 1 | | 37.71 | 12 | |
| 13 | 11 | | 15.93 | 13 | |
| 14 | 19 | | 36.53 | 14 | |
| 15 | 35 | | 25.62 | 15 | 0.131 |
| 16 | 30 | | 17.25 | 16 | |
| 17 | 12 | | 12.80 | 17 | |
| 18 | 34 | | 8.46 | 18 | |
| 19 | 14 | | 12.31 | 19 | |
| 20 | 24 | | 11.24 | 20 | 0.133 |
| 21 | 31 | | 5.61 | 21 | |
| 22 | 27 | | 5.77 | 22 | |
| 23 | 9 | | 29.47 | 23 | |
| 24 | 40 | | 9.53 | 24 | |
| 25 | 15 | | 4.68 | 25 | |
| 26 | | 14 | 0.01 | 24 | |
| 27 | 33 | | 17.73 | 25 | |
| 28 | 29 | | 5.15 | 26 | |
| 29 | 16 | | 2.47 | 27 | |
| 30 | | 1 | 0.01 | 26 | 0.122 |
| 31 | 32 | | 0.60 | 27 | |
| 32 | 22 | | 23.63 | 28 | |
| 33 | 6 | | 0.95 | 29 | |
| 34 | 1 | | 0.72 | 30 | |
| 35 | 38 | | 0.10 | 31 | |
| 36 | 17 | | 0.06 | 32 | |
| 37 | 28 | | 0.06 | 33 | |
| 38 | 18 | | 0.05 | 34 | 0.118 |
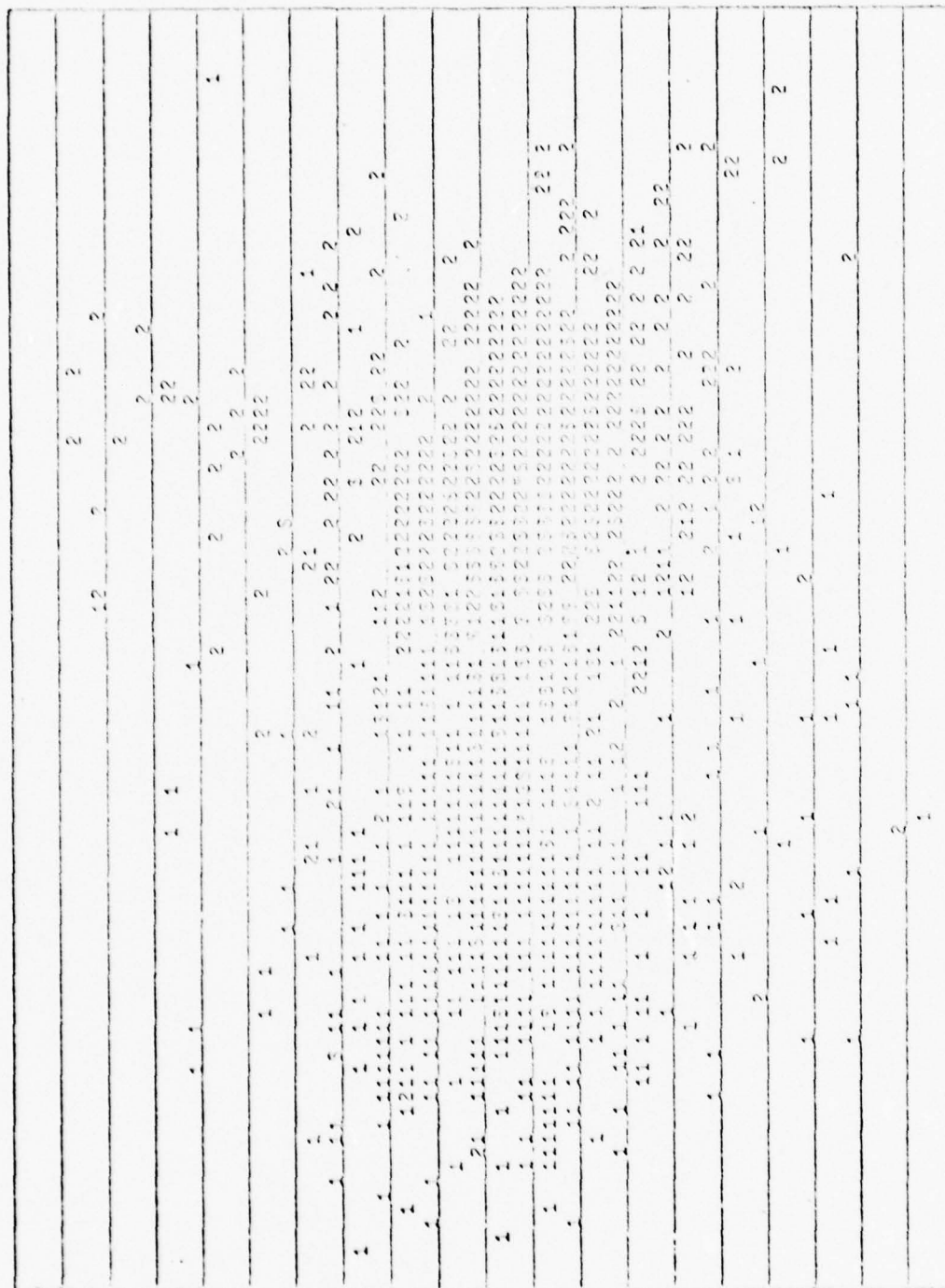
Fig 50. Scatterplot for Experiment 9

Table XI.  Features Selected at Various Steps for
            Experiment 9

| STEP NUMBER | FEATURES INCLUDED, ORDERED ON F |
|---|---|
| 1 | 36 |
| 5 | 36,25,39,37,8 |
| 10 | 36,25,5,39,37,26,2,7,8,3 |
| 20 | 36,39,25,5,11,3,19,37,1,2,26,34,35, 4,7,14,24,30,12 |
| 30 | 36,39,25,30,27,37,26,9,12,2,15, 33,35,5,40,11,31,19,3,34,29,7,24, 8,16,4 |
| 38 | 36,39,25,37,30,32,29,22,9,19,12, 27,15,26,11,33,5,40,2,3,35,7,16, 8,24,4,31,6,34,1,38,28,17,18 |

Table XII.  Confusion Matrix for Experiment 9

|  |  | CLASSIFIED AS | |
|---|---|---|---|
|  |  | 1 | 2 |
| TRUE | 1 | 698/782 | 136/52 |
| CLASS | 2 | 65/55 | 838/848 |

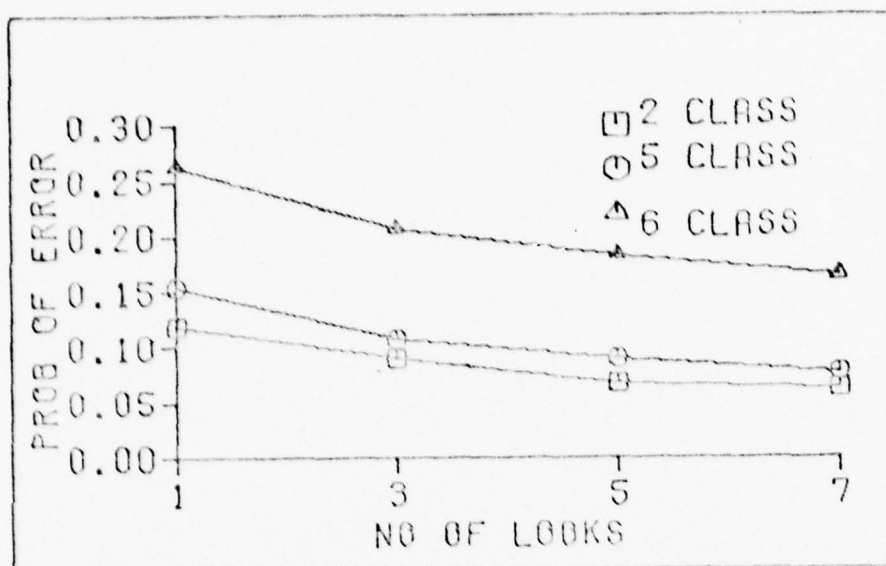Fig 31. Expanded Data Base Performance

Table XIII. Features Selected at Various Steps for
Experiment 10

| STEP NUMBER | FEATURES INCLUDED, ORDERED ON F |
|---|---|
| 1 | 15 |
| 5 | 5,2,3,15,39 |
| 10 | 36,5,39,3,37,12,28,2,15,18 |
| 20 | 36,37,39,31,38,35,5,20,32,28,22,27,3,2, 15,18,12,8,33,23 |
| 35 | 36,39,38,35,37,19,11,3,29,25,9, 4,27,12,1,30,22,31,5,40,32,18,28,17, 26,2,33,15,6,8,20,24,16,23,34 |

represented. The probability of error for Experiment 10 is plotted vs number of observations in Fig 31, along with the results considering Target 2 and 3 belonging to the same class. The six class probability of error for seven looks was .164. The Experiment 10 scatter plot is shown in Fig 32.

Table XIV. Experiment 10 Confusion Matrix

|  |  | CLASSIFIED | AS |  |  |  |  |
|---|---|---|---|---|---|---|---|
|  |  | 1 | 2 | 3 | 4 | 5 | 6 |
|  | 1 | 600 | 31 | 56 | 33 | 18 | 13 |
| TRUE | 2 | 11 | 228 | 107 | 2 | 1 | 0 |
| CLASS | 3 | 13 | 79 | 378 | 0 | 0 | 1 |
|  | 4 | 11 | 15 | 4 | 51 | 13 | 0 |
|  | 5 | 3 | 0 | 2 | 0 | 13 | 0 |
|  | 6 | 12 | 19 | 12 | 1 | 0 | 21 |

In the final experiment, number 11, the same data base as in Experiment 10 was used. In this experiment a tree structure was imposed on the classifier as suggested by Meisel (Refs 43 and 53). First, the data was classified into the two major groups as in Experiment 9 and then final classification was made. Since no significant difference in performance over Experiment 9 resulted, the results will not be detailed. The scatter plots of the two major groups are of interest and are shown in Fig 33 and Fig 34. The similiarity between targets 2 and 3 is manifest in Fig 34.
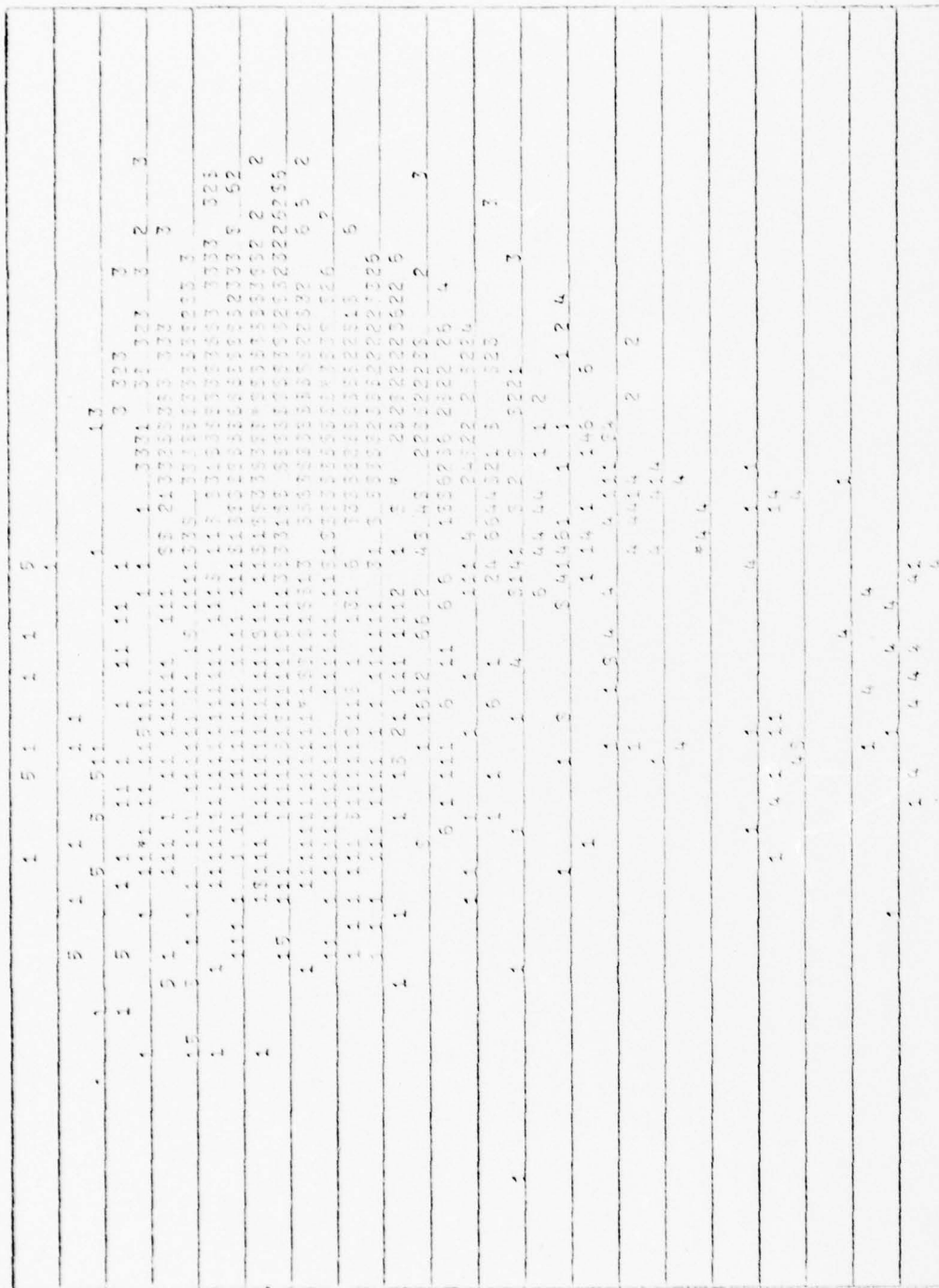
Fig 32. Scatterplot for Experiment 10
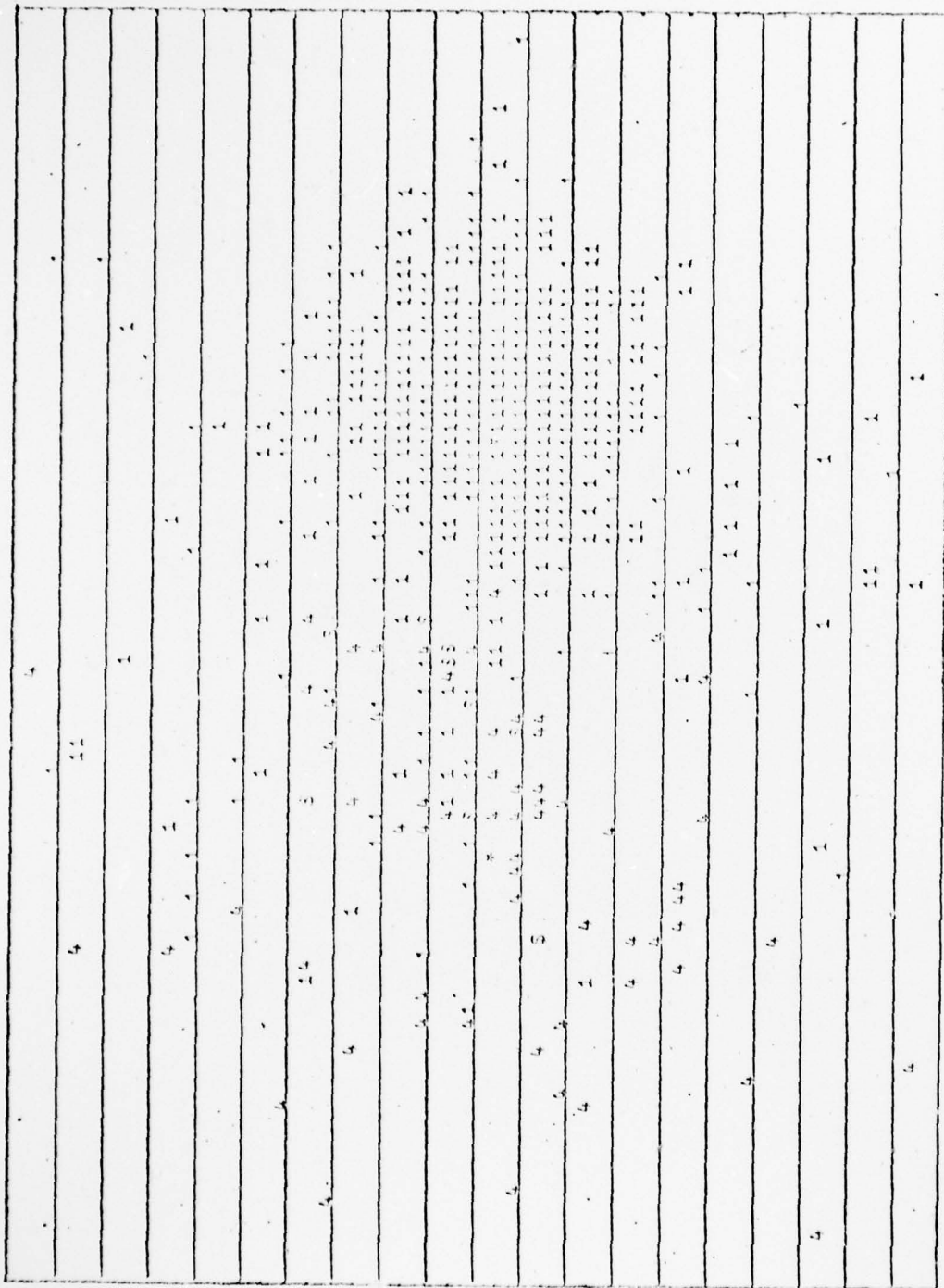
Fig 33. Scatterplot of Target 1 vs Target 4
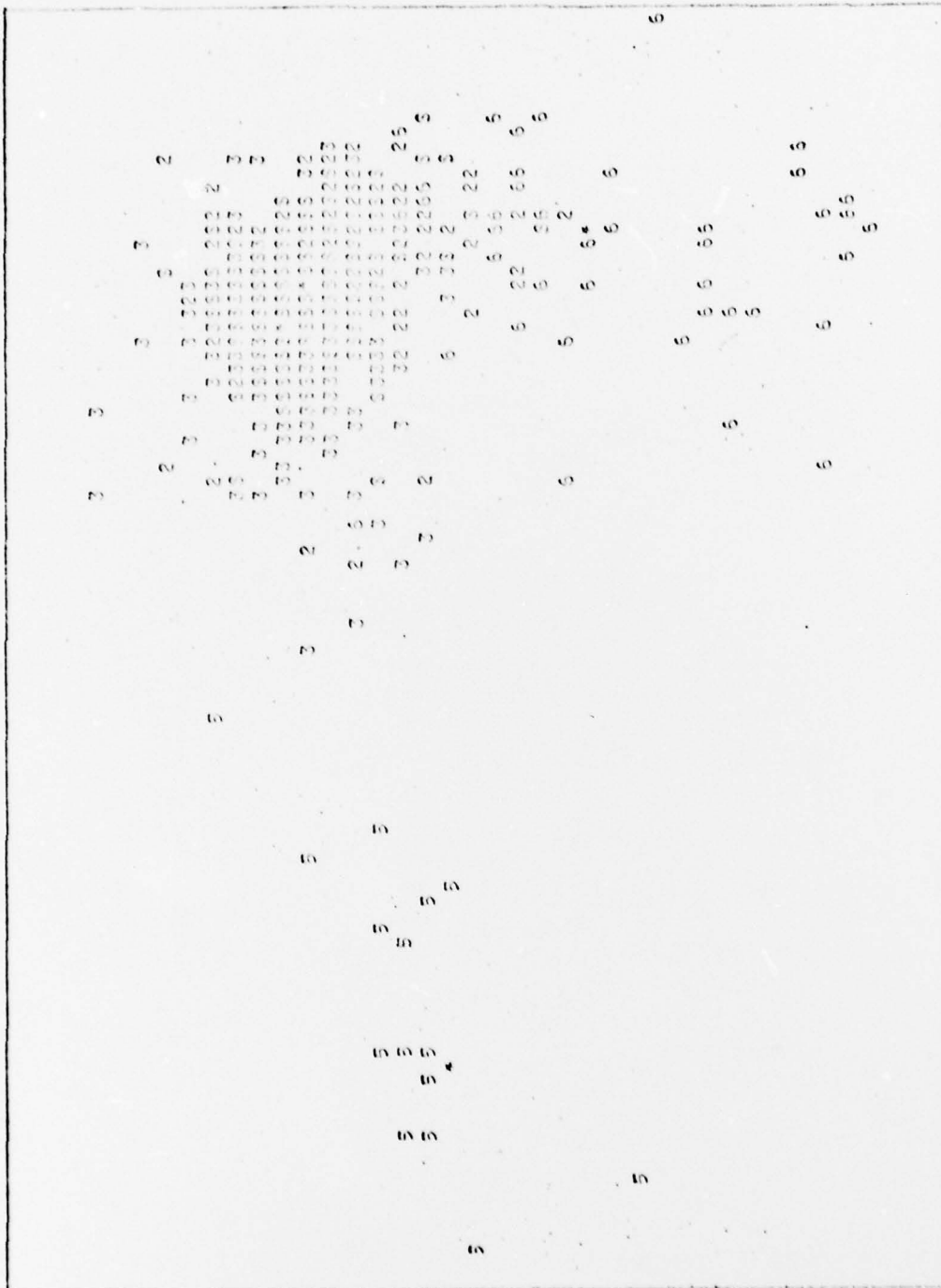
Fig. 54. Scatterplot of Targets 2, 3, 5, and 6

# V  Conclusion

This concluding chapter presents a summary of the work performed, conclusions to be drawn, and recommendations for further related research.

## Summary

This dissertation has presented a method of performing automatic recognition for the class of radar targets that possess periodically amplitude modulated signatures. The field of automatic target identification was surveyed, with particular attention given to radar. A philosophical discussion of pattern recognition followed by a procedure for designing pattern recognition systems was presented. The nature of the periodically amplitude modulated phenomenon was investigated, and the data bases used for experimentation were described. The form of the optimal classifier for this type of problem was examined, but it was concluded that such a classifier was not realizable due to inadequacies in the physical model and lack of adequate statistical information. Thus, a suboptimal frequency-domain classification algorithm was designed. Features that were relatively invariant with respect to target parameters were heuristically extracted from the short-time amplitude spectra of target signatures. Performance was presented for both linear discriminant and nearest neighbor classifiers.

## Conclusions

The periodically amplitude modulated signature model of vehicular radar targets introduced here appears to qualitatively account for the observed doppler spread. The model fits best at azimuth aspects of zero and 180 degrees, where the specular scattering from the modulating structure is greatest. The modulation is essentially frequency band-limited to the band between the carrier frequency and twice the target doppler frequency. The implication is that information only out to the

112

second harmonic of the skin line need be retained for classification purposes.

Based on both physical considerations and empirical observations, the obvious choice for the pattern space is the space of short-time Fourier amplitude spectra. The effects of the amplitude modulation are much more manifest in the frequency domain than in the time domain. Also, since the signal is essentially band-limited, it tends not to be time-limited. With the advent of FFT processors on an IC chip, the required preprocessing can be done quickly and cheaply. For the type of features used in this work, the 1024 point FFT with Hanning window provided sufficient resolution and stability.

Probabilities of error of under ten percent were achieved on five and fewer class problems by taking more than one observations. Table XV summarizes the classification results obtained for selected experiments. The most reliable overall feature is F(36) which is a measure of the spread of the spectrum about the skin line. Except for Experiment 5, typically about 35 features were used for estimating performance in Table XV, but performance should not be significantly affected by using only ten. Since there is no significant difference in performance between NN and SDA classifiers, the data for each class is essentially unimodal in the feature space. The very slight edge in performance of the 3NN over the NN classifier does not justify the increased computational expense. Although the cepstrum was not judged useful at all aspect angles, it could be used at zero and 180 degree aspect angles when the spectral features are less reliable. In all experiments, performance improved considerably as increased number of observations were considered. The decrease in probability of error tended to flatten out at about seven looks.

Table XV. Summary of Classification Results

| Exp. | Data Base | No. Classes | Classifier | Best Feature | Best 10 Features | Mean P(e) 1 Look | 7 Looks |
|---|---|---|---|---|---|---|---|
| 1 | A | 2 | SDA | 28 | 36,25,38,26,21,29, 28,34,37,23 | .061 | .026 |
| 3 | A | 2 | NN | | | .097 | .029 |
| 4 | A | 2 | 3NN | | | .089 | .021 |
| 5 | A | 2 | SDA | 10 | | .128 | .067 |
| 6 | A | 3 | SDA | 36 | 36,37,8,4,39,9,6, 12,14,22 | .219 | .150 |
| 7 | A | 3 | NN | | | .208 | .126 |
| 8 | A | 3 | 3NN | | | .203 | .133 |
| 9 | A&B | 2 | SDA | 36 | 36,25,5,39,37,26,2, 7,8,3 | .118 | .062 |
| 10 | A&B | 5 | SDA | | | .153 | .077 |
| 10 | A&B | 6 | SDA | 15 | 36,5,39,3,37,12,28, 2,15,18 | .263 | .164 |

Experiment 5 indicated that the techniques outlined here could possibly be extended to airborne radar with some success.

The results of this research indicate that it would be feasible to identify PAM targets by major type using the techniques applied. Classification of many types of targets into specific classes would require the use of more of the fine structure of the spectral signatures. Such an effort would require a more careful study of the sources of the periodic modulation and the noise sources.

## Recommendations for Further Research

No claims of optimality are made for the techniques proposed in

114

this thesis. Because of the nonstationarity of the random processes of interest and the lack of a complete physical model, an optimal solution does not appear likely in the near future. Since heuristic procedures are called for, there may well be approaches to the problem that will yield better results. It is obvious that further physical investigations into the nature of the amplitude modulation and the sources of noise would produce a more complete model.

An extension of the techniques proposed here, would be a sensitivity analysis to determine the optimal order of FFT to be used. The 1024 point FFT used in this work yields a theoretical frequency resolution of about two Hz and an actual resolution nearer four Hz when the finite window effect is considered. A lower order FFT would sacrifice some resolution but would produce a smoother spectrum. Also, the shorter the FFT integration period, the more nearly stationary the process would be during the period, resulting in less smearing of the short-time spectrum. Two other side benefits of a lower order FFT would be reduced computational requirement for preprocessing and feature extraction, and a larger design set. If an experiment was conducted to calculate performance in terms of average probability of error as a function of FFT order, performance might improve as FFT order was reduced from 1024 to a certain order, and then it would begin to deteriorate as the frequency resolution becomes too poor.

Another experiment that would be of interest would be to estimate the short-time spectra of the processes using the autoregressive spectral estimator rather than the FFT. Kaveh (Ref 40) has used this technique for high resolution velocity estimation of radar targets in the presence of extended clutter. Autoregressive (linear predictive) spectral

estimation is also widely used in speech analysis (Ref 43). The periodically amplitude modulated radar waveform bears a striking similarity to the speech signal. The order of the autoregressive estimator to be used could be determined by considering the maximum number of spikes that would be expected in the spectrum. The resulting spectrum would be a smoothed version, and no computational effort would be wasted, as for the FFT, in computing the amplitude values of frequency components that are not of interest.

The final recommendation for further research is to determine how significantly the identification performance would be degraded by distributed ground clutter as encountered in an airborne radar. This could be done by either gathering airborne data or by developing a suitable ground clutter model that could be added to the existing data. A model for distributed clutter that might be of use has been proposed by Ringel (Ref 56). If the identification procedure can be extended to the airborne problem, further work would be required to relate probability of error to aspect angle. Such a study might suggest optimum flight paths for aircraft engaged in an identification task.

# Bibliography

1.  Andrews, Harry C. Introduction to Mathematical Techniques in Pattern Recognition. New York: John Wiley and Sons, Inc., 1972.

2.  Becker, Peter W. Recognition of Patterns Using the Frequencies of Occurence of Binary Words. New York: Springer-Verlag, 1974.

3.  Bisignani, W. T. Automatic Target Classification. Technical Report MTR-6399. Washington, D.C.: MITRE Corporation, April 1973.(AD 763 493).

4.  Blackman, R. B. and J. W. Tukey. The Measurement of Power Spectra. New York: Dover Publications, Inc., 1959.

5.  Bogert,M. J. R., et al. "The Quefrency Alanysis of Time Series for Echoes: Cepstrum, Pseudo-autocovariance, Cross-cepstrum and Saphe Cracking," Proceedings of the Symposium on Time Series Analysis, edited by M. Rosenblatt. New York: John Wiley and Sons, Inc., 1963.

6.  Bouvier, Ronald D. Seismic Pattern Recognition. MS Thesis. Wright-Patterson AFB, Ohio: Air Force Institute of Technology, December 1972. (AD 757 877).

7.  "Business Today: Thinking Small," Time, 111:50-51 (February 1978).

8.  Chandrasekaran, B. and Anil K. Jain. "Independence, Measurement Complexity, and Classification Performance," IEEE Transactions on Systems, Man, and Cybernetics. SMC-5: 240-244 (March 1975).

9.  Chang, C. Y. "Dynamic Programming as Applied to Feature Subset Selection in Pattern Recognition Systems," IEEE Transactions on Systems, Man, and Cybernetics, SMC-3: 166-171 (March 1973).

10. Childers, Donald G, et al. "The Cepstrum: a Guide to Processing," Proceedings of the IEEE, 65: 1428-1443 (October 1977).

11. Chuang, C. W. and A. A. Ksienski.Backscatter of a Large Rotating Conducting Cylinder with Arbitrary Cross Section. Technical Report. 783815-4. Columbus, Ohio: Electroscience Laboratory, Ohio State University, February, 1978.

12. Cooley, J. W. and J. W. Tukey, "An Algorithm for the Machine Calculation of Complex Fourier Series," Mathematics of Computation, 19: 297-301 (April 1965).

13. Cover, Thomas M. "Review of Five Books in Pattern Recognition," IEEE Transactions on Information Theory, IT-19: 827-833 (November 1973). (AD A 023 376).

14. ------. "The Best Two Independent Measurements Are Not the Two Best," IEEE Transactions on Systems, Man, and Cybernetics, SMC-4: 116-117 (January 1974).

15. ----- and Peter E. Hart. "Nearest Neighbor Pattern Classification," IEEE Transactions on Information Theory, IT-13: 21-27 (January 1967).

16. ----- and Jan M. Van Campenhout. "On the Possible Orderings in the Measurement Selection Problem," IEEE Transactions on Systems, Man, and Cybernetics, SMC-7: 657-661 (September 1977).

17. Dixon, W. J., ed. BMD Biomedical Computer Programs. Berkeley: University of California Press, 1970.

18. Duda, Richard O. and Peter E. Hart. Pattern Classification and Scene Analysis. New York: John Wiley and Sons, Inc., 1973.

19. Dunn, John H. and Dean D. Howard. "Target Noise," Radar Handbook, edited by Merrill I. Skolnik. New York: McGraw-Hill Book Co., 1970.

20. Fishbein, William. Radar Division Chief, US Army Combat Surveillance and Target Acquisition Laboratory (personal correspondence). Fort Monmouth, New Jersey, February 15, 1978.

21. Fisher, R. A. "The Use of Multiple Measurements in Taxonomic Problems," Annals of Eugenics, 7: 179-188 (1936).

22. Fix, Evelyn and J.L. Hodges. Discriminatory Analysis: Nonparametric Discrimination. Project 21-49-004, Report 4. Randolph AFB, Texas: USAF School of Aviation Medicine, February 1951.

23. Foley, Donald H. "Considerations of Sample and Feature Size," IEEE Transactions on Information Theory, IT-18: 618-626 (September 1972).

24. Fu, King Sun. Syntactic Methods in Pattern Recognition. New York: Academic Press, 1974.

25. Fukunaga, Keinosuke. Introduction to Statistical Pattern Recognition. New York: Academic Press, 1972.

26. ----- and Pattrenahalli M. Narendra. "A Branch and Bound Algorithm for Feature Subset Selection," IEEE Transactions on Computers, C-26: 917-922 (September 1977).

27. Gersch, Will and James M. Kennedy. "Spectral Measurements of Sliding Tones," IRE Convention Record, Part 2: 157-170, 1960.

28. Gobien, Jurgen O. Simultaneous Detection and Estimation: the Use of Sufficient Statistics and Reproducing Probability Densities. Ph D Dissertation. Ann Arbor, Michigan: University of Michigan, November 1973 (AD 770 176).

29. Goggins, William B. Identification of Radar Targets by Pattern Recognition. Ph D Dissertation. Wright-Patterson AFB, Ohio: Air Force Institute of Technology, February 1973 (AD 764 713).

30. Goodman, Joseph W. Introduction to Fourier Optics. San Francisco:

McGraw-Hill Book Co., 1968.

31. Grenander, U. and G. Szego. Toeplitz Forms and Their Applications. Los Angeles: University of California Press, 1958.

32. Harris, Frederic J. "On the Use of Windows for Harmonic Analysis with the Discrete Fourier Transform," Proceedings of the IEEE, 66: 51-83 (January 1978).

33. Hart, Peter E. "The Condensed Nearest Neighbor Rule," IEEE Transactions on Information Theory, IT-14: 515-516 (May 1968).

34. Integrated Systems Software Corporation. DISSPLA, Vol I and II. San Diego, 1970.

35. International Mathematical and Statistical Libraries, Inc. IMSL Reference Manual. Houston, 1977.

36. Jennrich, Robert I. "Stepwise Discriminant Analysis," Statistical Methods for Digital Computers, edited by Kurt Enslein, et al. New York: John Wiley and Sons, Inc., 1977.

37. Kabrisky, Matthew and Joseph W. Carl. "Sequency Filtered Densely Connected Transforms for Pattern Recognition," Proceedings of the Fourth Hawaii International Conference on Systems Science. 215-220. University of Hawaii, January 1971. (AD 761 506).

38. Kaiser, J. F. "Digital Filters," System Analysis by Digital Computer, edited by F. F. Kuo and J. F. Kaiser. New York: John Wiley and Sons, Inc., 1966.

39. Kanal, Laveen. "Patterns in Pattern Recognition: 1968-1974," IEEE Transactions on Information Theory, IT-20: 697-722 (November 1974).

40. Kaveh, Mostafa. High Resolution Velocity Estimation in the Presence of Extended Clutter. PhD Dissertation. Purdue University, December 1974.

41. Ksienski, Aharon A., et al. "Low-Frequency Approach to Target Identification," Proceedings of the IEEE, 63: 1651-1660 (December 1975).

42. Luenberger, David G. Optimization by Vector Space Methods. New York: John Wiley and Sons, Inc., 1969.

43. Markel, J. D. and A. H. Gray. Linear Prediction of Speech. New York: Springer-Verlag, 1976.

44. Meisel, William S. Computer-Oriented Approaches to Pattern Recognition. New York: Academic Press, 1972.

45. Mohn, William S., Jr. "Two Statistical Feature Selection Techniques Applied to Speaker Identification," IEEE Transactions on Computers, C-20: 979-987 (September 1971).

46. Mucciardi, Anthony N. and Earl E. Gose. "A Comparison of Seven Techniques for Choosing Subsets of Pattern Recognition Properties," IEEE Transactions on Computers, C-20: 1023-1031 (September 1971).

47. Nahin, Paul J. "IFFN, a Technological Challenge for the '80s," Air University Review, 28: 1-16 (September 1977).

48. Nichol, D. G. "Processing Noisy Line Spectrograms as Digital Pictures," Pattern Recognition, 9: 137-146 (October 1977).

49. Nilsson, Nils J. Learning Machines. New York: McGraw-Hill Book Co., 1965.

50. Ohio State University Department of Electrical Engineering. Radar Target Identification. Short Course Class Notes. Columbus, Ohio: Ohio State University, September 1977.

51. Oppenheim, Alan V. and Ronald W. Schafer. Digital Signal Processing. Englewood Cliffs, New Jersey: Prentice Hall, Inc., 1975.

52. Pau, L. F. and M. Delosme. "Classification of Simple Objects Using Laser Rangefinder Measurements," Proceedings of the Third International Joint Conference on Pattern Recognition. 783-788. Coronado, California: IEEE, November 1976.

53. Pavlidis, Theodosios. A Linguistic Model for Waveform Analysis. Engineering Technical Report. Princeton, New Jersey: Department of Electrical Engineering, Princeton University, March 1969. (AD 684 360).

54. Payne, Harold J. and William S. Meisel. "An Algorithm for Constructing Optimal Binary Decision Trees," IEEE Transactions on Computers, C-26: 905-916 (September 1977).

55. Radov, C. Pattern Recognition by Fourier Series Transformation. MS Thesis. Wright-Patterson AFB, Ohio: Air Force Institute of Technology, March 1962.

56. Ringel, Melvin B. "An Advanced Computer Calculation of Ground Clutter in an Airborne Pulse Doppler Radar," Proceedings of the National Aerospace and Electronics Conference. 921-928. Dayton, Ohio: IEEE, May 1977.

57. Robinson, Prentiss N. and Jane B. DeNuzzo. "Target Spectra Feature Selection Using A Priori Information," IEEE Transactions on Aerospace and Electronic Systems, AES-11: 650-653 (July 1975).

58. Rosenbaum, F. and E. G. Klimchak. Target (Tank) Recognizer. Technical Report. Towson, Maryland: Bendix Corporation, August 1974. (AD 922 333).

59. Rosenfeld, Azriel and Avinash C. Kak. Digital Picture Processing. New York: Academic Press, 1976.

60. Shannon, C. E. "A Mathematical Theory of Communication," Bell System Technical Journal, 27: 379-423 (July 1948).

61. Shrihari, S. N. Comparative Evaluation of the Sebestyen and Nearest Neighbor Classifiers for Radar Aircraft Identification, PhD dissertation. Columbus, Ohio: Ohio State University, June 1976. (AD A 029 549).

62. Singleton, Richard C. "On Computing the Fast Fourier Transform," Communication of the ACM, 10: 647-654 (October 1967).

63. Skolnik, Merrill I. Introduction to Radar Systems. New York: McGraw-Hill Book Co., 1962.

64. Stearns, Stephen D. "On Selecting Features for Pattern Classifiers, "Proceedings of the Third International Joint Conference on Pattern Recognition: 71-75. Coronado, California IEEE, November 1976.

65. Thomas, D. W. and B. R. Wilkins. "The Analysis of Vehicle Sounds for Recognition," Pattern Recognition, 4: 379-389 (December 1972).

66. Toussaint, Godfried T. "Bibliography on Estimation of Misclassification." IEEE Transactions on Information Theory, IT-20: 472-479 (July 1974).

67. Ullman, J. R. Pattern Recognition Techniques. London: Butterworth and Co., 1973.

68. Van Bladel, Jean. "Electromagnetic Fields in the Presence of Rotating Bodies," Proceedings of the IEEE, 64: 301-318 (March 1976).

69. Van Trees, Harry L. Detection, Estimation, and Modulation Theory, Part I. New York: John Wiley and Sons, Inc., 1968.

70. ------. Detection, Estimation, and Modulation Theory, Part III. New York: John Wiley and Sons, Inc., 1971.

71. Wang, N. "Self-Consistent GTD Formulation for Conducting Cylinders with Arbitrary Convex Cross Section," IEEE Transactions on Antennas and Propagation, AP-24: 463-472 (March 1976).

72. Zadeh, Lofti A. "Fuzzy Sets," Information and Control, 8: 338-358, 1965.

VITA

Clayton Verne Stewart was born on 3 October 1942 in Roswell, New Mexico. He graduated from high school in Lubbock, Texas in 1960 and attended Texas Technological College before entering the Air Force Aviation Cadet program in June 1963. Upon completion of flight training in July 1964 he was awarded his commission and navigator rating and was assigned to Electronic Warfare Officer (EWO) training. His subsequent assignments included EWO duties aboard EB-57, EB-66, and C-130 aircraft in Air Defense Command, Pacific Air Forces, Tactical Air Command, and Headquarters Command. In January 1973 he completed the requirements for the Bachelor of Science degree in Engineering Science at the University of Redlands, and in June 1974 he entered the School of Engineering, Air Force Institute of Technology. He received the MSEE degree and was awarded the Mervin E. Gross award for academic excellence in December 1975. He is currently serving as a Major in the Air Force and is a doctoral candidate at the Air Force Institute of Technology. He is a member of IEEE, the Pattern Recognition Society, the Association of Old Crows, Eta Kappa Nu, and Tau Beta Pi.

Permanent address:   Department of Electrical Engineering

US Air Force Academy

Colorado  80840

122

| REPORT DOCUMENTATION PAGE | | READ INSTRUCTIONS<br>BEFORE COMPLETING FORM |
|---|---|---|
| 1. REPORT NUMBER<br>AFIT/DS/EE/78-2 | 2. GOVT ACCESSION NO. | 3. RECIPIENT'S CATALOG NUMBER |
| 4. TITLE (and Subtitle)<br>IDENTIFICATION OF PERIODICALLY AMPLITUDE<br>MODULATED TARGETS | | 5. TYPE OF REPORT & PERIOD COVERED<br>Ph D Dissertation |
| | | 6. PERFORMING ORG. REPORT NUMBER |
| 7. AUTHOR(s)<br>Maj Clayton V. Stewart | | 8. CONTRACT OR GRANT NUMBER(s) |
| 9. PERFORMING ORGANIZATION NAME AND ADDRESS<br>Air Force Institute of Technology (AFIT-EN)<br>Wright-Patterson AFB, Ohio 45433 | | 10. PROGRAM ELEMENT, PROJECT, TASK<br>AREA & WORK UNIT NUMBERS<br>Project 7622-01-31 |
| 11. CONTROLLING OFFICE NAME AND ADDRESS<br>Radar Branch (AFAL-RWM)<br>Air Force Avionics Laboratory<br>Wright Patterson AFB, Ohio 45433 | | 12. REPORT DATE<br>May 1978 |
| | | 13. NUMBER OF PAGES<br>134 |
| 14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office) | | 15. SECURITY CLASS. (of this report)<br>Unclassified |
| | | 15a. DECLASSIFICATION DOWNGRADING<br>SCHEDULE |

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

Approved for public release; IAW AFR 190-17

JERRAL F. GUESS, Captain, USAF
Director of Information

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Fast Fourier Transform          Signal Processing
Feature Extraction
Pattern Recognition
Radar
Target Identification

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

An algorithm that may be used for the classification of periodically ampli-
tude modulated (PAM) targets is presented. The data base used to test the algo-
rithm is derived from radar returns from vehicles moving at various velocities
and aspect angles, but the techniques are applicable, as well, to other active
wave devices such as sonar and laser. The received radar signal is considered
to be a time series that is a function of target type, range, velocity, orien-
tation, and noise. Classification is implemented in the frequency domain; short-

DD FORM 1473 EDITION OF 1 NOV 65 IS OBSOLETE

(20) time spectra are computed using the Fast Fourier Transform (FFT). Features are extracted from the information-bearing sidebands of the resulting spectra. The radar signatures are classified using both linear discriminant and nearest neighbor classifiers, and performance is presented for two, three, five and six class cases using single and sequential looks. Probabilities of error of less than ten percent are achieved for five or fewer classes.